Generalized Least Squares-Based Parametric Motion Estimation Under Non-uniform Illumination Changes*

Raúl Montoliu[1] and Filiberto Pla[2]

[1] Dept. Arquitectura y Ciéncias de los Computadores.
[2] Dept. Lenguajes y Sistemas Informáticos.
Jaume I University
Campus Riu Sec s/n 12071 Castellón, Spain
{montoliu,pla}@uji.es
http://www.vision.uji.es

Abstract. The estimation of parametric global motion has had a significant attention during the last two decades, but despite the great efforts invested, there are still open issues. One of the most important ones is related to the ability to recover large deformation between images in the presence of illumination changes while kipping accurate estimates. In this paper, a Generalized least squared-based motion estimator is used in combination with a dynamic image model where the illumination factors are functions of the localization (x, y) instead of constants, allowing for a more general and accurate image model. Experiments using challenging images have been performed showing that the combination of both techniques is feasible and provides accurate estimates of the motion parameters even in the presence of strong illumination changes between the images.

1 Introduction

The estimation of motion in images is a basic task in computer vision with many application fields. One of the most important goals in the motion estimation field is to estimate the motion as accurately as possible. The problem of motion estimation is not an easy task when there are large deformations and illumination variations between images. In addition, the presence of areas which do not support the main motion (outliers) is an additional source of inaccuracy.

Traditionally, the motion estimation problem has been formulated following the assumption that the changes in gray levels between images are only due to motion, i.e. the Brightness Constancy Assumption (*BCA*). However, a pixel can change its brightness value because an object moves to another part of the scene with different illumination or because the illumination of the scene changes, locally or globally, between images. In these cases, the *BCA* fails, and therefore, it is not possible to obtain accurate estimates. To overcome this problem, two are

* This work has been partially funded with projects ESP2005-07724-C05 and CSD2007-00018 from the Spanish Ministry of Science and Education.

A. Campilho and M. Kamel (Eds.): ICIAR 2008, LNCS 5112, pp. 660–669, 2008.

the most commonly used techniques. On one hand, the images can be preprocessed to transform them to a new color space where shadows, highlights and other illumination effects have been partially removed [2,3,8]. Then, the motion estimator is applied to the modified images. Alternatively, a more complex image model than the *BCA* can directly be used in the motion estimation process. Thus, the estimator can calculate, at the same time, the motion and illumination parameters [4,5]. The second type of approach is the technique that has been used in this work. In particular, a dynamic image model where the multiplication and bias illumination factors are functions of the localization (x, y) instead of constants, has been used at this paper. This dynamic image model has been combined with a Generalized Least Squares-based (*GLS*) motion estimator [7] which obtains accurate estimates even when there exist large deformations between images and in the presence of an important number of outliers. Therefore, with the combination of both techniques (the GLS-based motion estimator and the use of the dynamic image model) a motion estimator can be obtained which can perform the motion estimation task in an accurate manner while allowing large deformation and non-uniform illumination changes between images. Thus, the main objective of this paper is to reformulate a motion estimator that use a constant illumination model to accommodate it to non-uniform illumination changes by using a spatial dynamic image model.

The rest of the paper is organized as follows: next section explains the dynamic image model used in this paper, Section 3 comments how the dynamic image model has been combined with the GLS-based motion estimator. Section 4 shows the main results and finally in the last section, the main conclusions drawn from this work are summarized.

2 Spatially Varying Illumination Model

Conventional intensity-based motion estimation methods are based on the brightness constancy assumption given as follows:

$$I_1(x_i, y_i) - I_2(x'_i, y'_i) = 0, (\forall i \in \Re), \tag{1}$$

where $I_1(x_i, y_i)$ is the gray level of the first image in the sequence (test image) at the point (x_i, y_i), and $I_2(x'_i, y'_i)$ is the gray level of the second image in the sequence (reference image) at the transformed point (x'_i, y'_i). $\Re$ is the region of interest.

Some preliminary works [11] used an illumination model to account for uniform photometric variation as follows:

$$\alpha I_1(x_i, y_i) + \beta - I_2(x'_i, y'_i) = 0, \tag{2}$$

where the constant α and β are the illumination multiplication and bias factor, respectively. The main problem of that illumination model is that it cannot account for spatially varying illumination variations. To overcome this restriction, a more general dynamic image model [9] can be used where the multiplication and

bias factor are functions of localization, i.e. $\alpha \equiv \alpha(x_i, y_i)$ and $\beta \equiv \beta(x_i, y_i)$. Assuming that these two illumination factors are slowly varying functions of localization, they can be well approximated by low-order polynomials. For instance, $\alpha(x_i, y_i)$ and $\beta(x_i, y_i)$ can be expressed using bilinear and constant polynomials, respectively, as follows:

$$\begin{aligned} \alpha(x_i, y_i) &= \alpha_x x_i + \alpha_y y_i + \alpha_c \\ \beta(x_i, y_i) &= \beta_c \end{aligned} \tag{3}$$

Applying this dynamic image model, Eq. (2) can be expressed using Eq. (3) as follows:

$$\alpha(x_i, y_i) I_1(x_i, y_i) + \beta(x_i, y_i) - I_2(x'_i, y'_i) = 0. \tag{4}$$

3 GLS-Based Motion Estimation Under Varying Illumination

The GLS-based motion estimator is a non-linear motion estimation technique proposed in [7] as an alternative method to M-Estimators [1,10] and other robust techniques to deal with outliers in motion estimation scenarios. The way how the motion estimation problem has been formulated provides an additional constraint that helps the matching process using image gradient information, since it is well known that the areas with more information for motion estimation are the ones that have intensity variations like in the object edges of the image. This constrain can be interpreted as a weight for each observation, providing high values to the weights of the observations considered as inliers, i.e. the ones that support the motion model, and low values to the ones considered as outliers.

The GLS-based motion estimation problem can be expressed as follows (see [7] for details):

$$\text{minimize } [\Theta_\upsilon = \upsilon^t \upsilon] \text{ subject to } F(\chi, \lambda) = 0, \tag{5}$$

where:

- υ is a vector of r unknown residuals in the observation space, that is, $\upsilon = \lambda - \tilde{\lambda}$, where λ and $\tilde{\lambda}$ are the unperturbed and actually measured vector of observations, respectively.
- $\chi = (\chi^1, \ldots, \chi^p)^t$ is a vector of p parameters.
- λ is made up of r elements λ_i, $\lambda = (\lambda_1, \ldots, \lambda_r)^t$, each one being an observation vector with $n = 3$ components $\lambda_i = (x_i, y_i, I_1(x_i, y_i))$.
- $F(\chi, \lambda)$ is made up of r functions $F_i(\chi, \lambda_i)$, $F(\chi, \lambda) = (F_1(\chi, \lambda_1), \ldots, F_r(\chi, \lambda_r))^t$. These function can be non linear.

The solution of (5) can be addressed as an iterative optimization starting with an initial guess of the parameters $\widehat{\chi}(0)$. At each iteration j, the algorithm estimates $\widehat{\Delta\chi}(j)$ to update the parameters as follows:

$$\widehat{\chi}(j) = \widehat{\chi}(j-1) + \widehat{\Delta\chi}(j). \tag{6}$$

The process is stopped if the improvement $\widehat{\Delta\chi}(j)$ at iteration j is smaller than a user-specified resolution in the parameter space. The desired expression $\widehat{\Delta\chi}(j)$ is calculated as follows:

$$\widehat{\Delta\chi}(j) = (A^t Q A)^{-1} A^t Q E, \tag{7}$$

where $A = \partial F/\partial\chi$, $B = \partial F/\partial\lambda$ and $E = -F(\widehat{\chi}(j), \widehat{\lambda}(j))$. The matrix $Q = (BB^t)^{-1}$ has been introduced to simplify the notation. Eq. (7) can also be expressed in a more convenient way as follows:

$$\widehat{\Delta\chi}(j) = \left(\sum_{i=1\ldots r} N_i \right)^{-1} \left(\sum_{i=1\ldots r} R_i, \right), \tag{8}$$

where $N_i = A_i^t (B_i B_i^t)^{-1} A_i$ and $R_i = A_i^t (B_i B_i^t)^{-1} E_i$, with

$$\begin{aligned} B_i &= \left(\tfrac{\partial F_i(\widehat{\chi}(j-1),\lambda_i)}{\partial \lambda_i^1} \ \ldots \ \tfrac{\partial F_i(\widehat{\chi}(j-1),\lambda_i)}{\partial \lambda_i^n} \right)_{(1\times n)}, \\ A_i &= \left(\tfrac{\partial F_i(\widehat{\chi}(j-1),\lambda_i)}{\partial \chi^1} \ \ldots \ \tfrac{\partial F_i(\widehat{\chi}(j-1),\lambda_i)}{\partial \chi^p} \right)_{(1\times p)}, \\ E_i &= \left(-F_i(\widehat{\chi}(j-1), \lambda_i) \right)_{(1\times 1)}. \end{aligned} \tag{9}$$

The vector of parameters χ depends on the motion and illumination models used. In this paper, affine motion (6 parameters) and bilinear (3 parameters) and constant (1 parameter) polynomials for multiplication and bias factors (see Eq. (3)), respectively, has been used. Therefore, the vector of parameters is $\chi = (a_1, b_1, c_1, a_2, b_2, c_2, \alpha_x, \alpha_y, \alpha_c, \beta_c)^t$. The transformed coordinates (x_i', y_i') are related to the original ones (x_i, y_i) in affine motion as follows:

$$\begin{cases} x_i' = a_1 x_i + b_1 y_i + c_1 \\ y_i' = a_2 x_i + b_2 y_i + c_2 \end{cases} \tag{10}$$

In the original method, each $F_i(\chi, \lambda_i)$ was expressed as follows: $F_i(\chi, \lambda_i) = I_1(x_i, y_i) - I_2(x_i', y_i')$, i.e. the *BCA*. In this paper, a dynamic image model (DIM) which allows spatially varying illumination is used instead (see Eq. (4)). Therefore, each $F_i(\chi, \lambda_i)$ is expressed as follows:

$$F_i(\chi, \lambda_i) = \alpha(x_i, y_i) I_1(x_i, y_i) + \beta(x_i, y_i) - I_2(x_i', y_i'). \tag{11}$$

In order to calculate the matrices A_i, B_i and E_i, the partial derivatives of the function $F_i(\chi, \lambda_i)$ with respect to the parameters and with respect to the

observations must be worked out. The resulting A_i, B_i and E_i using affine motion are expressed as follows:

$$\begin{aligned} B_i^t &= \begin{pmatrix} \alpha_x I_1 + \alpha(x_i, y_i) I_1^x - a_1 I_2^x - a_2 I_2^y \\ \alpha_y I_1 + \alpha(x_i, y_i) I_1^y - b_1 I_2^x - b_2 I_2^y \\ \alpha(x_i, y_i) \end{pmatrix} \\ A_i &= (-x_i I_2^x, -y_i I_2^x, -I_2^x, -x_i I_2^y, -y_i I_2^y, -I_2^y, x_i I_1, y_i I_1, I_1, 1.0) \\ E_i &= -\left(\alpha(x_i, y_i) I_1(x_i, y_i) + \beta_c - I_2(x_i', y_i')\right) \end{aligned} \tag{12}$$

where I_1^x, I_1^y, I_2^x and I_2^y have been introduced to simplify notation as: $I_1^x \equiv I_1^x(x_i, y_i)$, $I_1^y \equiv I_1^y(x_i, y_i)$, $I_2^x \equiv I_2^x(x_i', y_i')$ and $I_2^y \equiv I_2^y(x_i', y_i')$, with $I_1^x(x_i, y_i)$ and $I_1^y(x_i, y_i)$ being the gradients of the test image at point (x_i, y_i); and with $I_2^x(x_i', y_i')$ and $I_2^y(x_i', y_i')$ being the gradients of the reference image at point (x_i', y_i').

In this formulation of the motion estimation problem, the expression $(B_i B_i^t)^{-1}$ plays the role of a weight providing high values (close to 1) when the gradient values in the reference and the test image are similar and low ones (close to 0) in the opposite case. That is one the main reasons why the GLS-based estimator provides accurate estimates of the parameters (see [7] for details).

3.1 Motion and Illumination Parameters Initialization

In many motion estimation problems where the deformation between images is quite large (e.g. large rotation, strong changes of scale, etc.), it is necessary to initialize the motion estimator using a good initial vector of motion parameters. For this purpose, first a feature-based method is used to obtain the initial vector of parameters that are not very far from the true solution. Using this initialization (i.e. $\widehat{\chi}(0)$), in the second step, the GLS-based motion estimator using the dynamic illumination model, is applied, which refines the estimation of the motion and illumination parameters up to the accuracy level desired by the user (see Eq. (6)).

At the first step, to cope with strong changes of scale, rotations, illumination changes and partially affine invariance, a SIFT-based technique [6] has been used to detect and describe interest points. For each interest point belonging to the first image a *K-NN* search strategy is performed to find the k-closest interest points at the second image. Finally, for estimating the first approximation of the motion parameters a random sampling technique is used [12]. Regarding the illumination parameters at $\widehat{\chi}(0)$, they have initially been set to: $\alpha_x = \alpha_y = \beta_c = 0$ and $\alpha_c = 1$.

4 Experimental Results

In this section, a set of motion estimation experiments are performed in order to test the accuracy of the proposed technique. In particular, the accuracy of the estimation in the case of using the Brightness Constancy Assumption (*BCA*, see

Fig. 1. Sample images (up to bottom, left to right) from *Boat*, *Bikes*, *Bark*, *Leuven* and *satellite* image sets

Eq. (1)) is compared with the case when the dynamic image model (*DIM*, see Eq. (4)) is used instead. To check the accuracy of the estimation, the normalized correlation coefficient (*Ncc*) similarity measure has been calculated using the pixels of the overlapped area of both images. The *Ncc* gives values from −1.0 (low similarity) to 1.0 (high similarity), and is expressed as follows:

$$Ncc(I_1, I_2) = \frac{\sum_{(x_i,y_i)\in\Re}[(\alpha_i I_1 + \beta_i - \mu_1)(I_2 - \mu_2)]}{\sqrt{\sum_{(x_i,y_i)\in\Re}(\alpha_i I_1 + \beta_i - \mu_1)^2 \sum_{(x_i,y_i)\in\Re}(I_2 - \mu_2)^2}}, \qquad (13)$$

where μ_1,μ_2 are the average of the gray level of both images, $\Re$ the overlapped area and I_1, I_2, α_i and β_i have been introduced to simplify notation as: $I_1 \equiv I_1(x_i, y_i)$, $I_2 \equiv I_2(x'_i, y'_i)$, $\alpha_i \equiv \alpha(x_i, y_i)$ and $\beta_i \equiv \beta(x_i, y_i)$.

A set of challenging sets of image pairs have been selected. They can be downloaded from Oxford's Visual Geometry Group web page [1] except for the last set that has been obtained from Internet. Oxford's ones present three main types of changes between images in 4 different sets of images; Blur: *Bikes* set, global illumination: *Leuven* set and zoom+rotation: *Bark* and *Boat* sets. The *satellite* set is a set of images from the same area but they have been captured at different times/days and therefore with different illumination conditions. Each image set has 6 different images. A sample of the images are showed at Figure 1. For each set, the 6 images have been combined in all possible pairs ($1 \leftrightarrow 2$, $1 \leftrightarrow 3$, ..., $1 \leftrightarrow 6$, $2 \leftrightarrow 3$, ..., $5 \leftrightarrow 6$).

To introduce a large illumination variation in the data, the second image of each image pair $I_1 \leftrightarrow I_2$ is modified multiplying it by a multiplier function. Two

[1] `http://www.robots.ox.ac.uk/~vgg/research/affine/index.html`

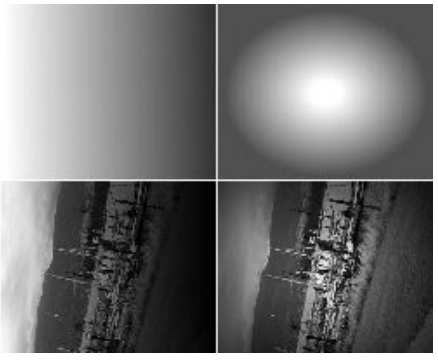

Fig. 2. The first row shows the multipliers used to add large illumination variation at data. The second row shows an example of the resulting images after applying the multipliers (the original image is the second one at the first row of Figure 1).

multipliers have been used, the first one makes dark the image from left to right and the second one has the form of a Gaussian. They are showed at the first row of Figure 2. The second row of Figure 2 shows an example of application of the multipliers. The resulting images, after the application of the multipliers, are called I_2^{Gd} and I_2^{Gn}, respectively. Note that the illumination changes of *Leuven* set are different from the ones introduced by the multipliers, since, in the first case, the changes are global, i.e. the changes do not depend on the localization of the pixel, while, in the second case, the multipliers introduces illumination changes which depend on the localization of the pixel.

For each image pair: $I_1 \leftrightarrow I_2$, the proposed motion estimation techniques is applied in order to obtain six Ncc values. First, the proposed motion estimation technique is performed using the original images (i.e. I_1 and I_2) with the BCA and the DIM to obtain two Ncc values: $Ncc(BCA)$ and $Ncc(DIM)$. In the second step, the image I_2^{Gd} is used as second image, producing the Ncc values: $Ncc^{Gd}(BCA)$ and $Ncc^{Gd}(DIM)$. Finally, the same process is repeated using now the image I_2^{Gn} obtaining the Ncc values: $Ncc^{Gn}(BCA)$ and $Ncc^{Gn}(DIM)$.

Tables 1 shows the median of the Ncc obtained for each set, when original image I_2 (i.e. image pair $I_1 \leftrightarrow I_2$), modified image I_2^{Gd} (i.e. image pair $I_1 \leftrightarrow I_2^{Gd}$) and modified image I_2^{Gn} (i.e. image pair $I_1 \leftrightarrow I_2^{Gn}$) are used as second image, respectively. The second and the third column show the median of the Ncc estimated for each image pair of each set, when using the BCA and DIM, respectively, as image model. The last column shows the percentage of times where the use of DIM improves the accuracy.

In general, the use of the dynamic image model instead of the BCA provides more accurate results in almost all the cases. Table 1(a) shows that although non additional illumination changes have been artificial added, the use of the dynamic image model improves the accuracy of the estimation, since, probably, there is a small (just no visually appreciable, but existing) illumination variation due to the acquisition process. The accuracy level is very similar in both cases, but

Table 1. Results obtained when the images I_2 (a), I_2^{Gd} (b) and I_2^{Gn} (c) are used as second image

(a) Image pairs $I_1 \leftrightarrow I_2$

Image Set	$Ncc(BCA)$	$Ncc(DIM)$	Best DIM
Bikes	0.9904	0.9907	93,33%
Bark	0.9636	0.9644	100.00%
Boat	0.9110	0.9238	100.00%
Leuven	0.9781	0.9796	100.00%
Satellite	0.9168	0.9203	100.00%

(b) Image pairs $I_1 \leftrightarrow I_2^{Gd}$

Image Set	$NCC^{Gd}(BCA)$	$NCC^{Gd}(DIM)$	Best DIM
Bikes	0.6785	0.9958	100.00%
Bark	0.4707	0.9748	100.00%
Boat	0.4248	0.9624	100.00%
Leuven	0.1809	0.9885	100.00%
Satellite	0.2987	0.9649	100.00%

(c) Image pairs $I_1 \leftrightarrow I_2^{Gn}$

Image Set	$NCC^{Gn}(BCA)$	$NCC^{Gn}(DIM)$	Best DIM
Bikes	0.4522	0.6476	100.00%
Bark	0.8683	0.9344	100.00%
Boat	0.5239	0.8455	100.00%
Leuven	0.3137	0.9055	100.00%
Satellite	0.3055	0.7385	100.00%

almost always the use of the DIM improve the accuracy of the estimation. Note that at *Leuven* set, the GLS-based motion estimator obtains accurate estimates even when the *BCA* is used, that is due to the weights used at the estimation procedure depend on gradient information and not on the grey level (see [7] for details).

Tables 1(b) and (c) show how the accuracy of the estimation is drastically reduced when using the *BCA*, since the strong illumination changes introduced make that the *BCA* is not fulfilled at the majority of the observations and, therefore, the estimation procedure gets lost while searching for the optimal parameters in the minimization process. The dynamic image model can deal with this situation, and therefore, when using it, the accuracy of the estimation is improved. Note how the accuracy level obtained when using the first multiplier (i.e. using I_2^{Gd} as the second image) is as good, and even better, as the cases when no illumination changes have been introduced to second image.

The second multiplier (i.e. using I_2^{Gn} as second image) introduces stronger illumination changes than the fist one, since the illumination changes introduced can not perfectly been modelled using the dynamic model proposed. Therefore, the accuracy obtained is not as good as the previous case, but it is still maintains high accuracy levels.

In order to show an illustrative example of the behavior of the proposed approach, a mosaic image has been created using the motion parameters obtained

Fig. 3. Mosaic image created using as input images the first image showed at Figure 1 as I_1, and the left image of the second row of Figure 2 as I_2

Fig. 4. Original I_1 (left), estimated I_1' (middle) and grey level difference (right)

from the motion estimation experiment which has as input images the image pair formed by the first image showed at Figure 1 as I_1, and the left image of the second row of Figure 2 as I_2. In despite of the strong illumination differences between both inputs images, the proposed techniques obtains accurate estimates as can be showed at Figure 3.

In order to see how the illumination parameters have also been accurately estimated, the second image of the previous experiment has been transformed using the motion and illumination parameters obtained with the proposed approach. That image (called I_1') should be quite similar to the first image of the pair if the parameters have been accurately estimated. Figure 4 show both images and the grey level differences between both images. In spite of both input images have been captured at different time moments, and therefore, the vegetation, people, water and even the boat are not completely stationary between both images, the images show that the illumination parameters have been estimated with high accuracy.

5 Conclusions

In this paper, an accurate Generalized least squared-based motion estimator is used in combination with a dynamic image model where the multiplication and bias illumination factors are functions of the localization (x, y). Experiments using challenging real images have been performed to show that with the combination of both techniques a motion estimator can be obtained which can perform the motion estimation task in an accurate manner while allowing large deformation and illumination changes between images.

References

1. Bober, M., Kittler, J.V.: Robust motion analysis. In: IEEE Conf. on Computer vision and Pattern Recognition, pp. 947–952 (1994)
2. Graham, D., Finlayson, S.D., Drew, M.S.: Removing shadows from images. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 823–836. Springer, Heidelberg (2002)
3. Geusebroek, J.M., van den Boomgaard, R., Smeulders, A.W.M., Geerts, H.: Color invariance. IEEE Trans. Pattern Anal. Machine Intell. 23(12), 1338–1350 (2001)
4. Kim, Y., Martinez, A.M., Kak, A.C.: Robust motion estimation under vaying illumination. Image and Vision Computing 23(4), 365–375 (2004)
5. Lai, S.-H., Fang, M.: Robust and efficient image alignment with spatially varying illumination models. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 02, p. 2167 (1999)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. Montoliu, R., Pla, F.: Generalized least squares-based parametric motion estimation. Technical Report 1/10/2007, University Jaume I (October 2007)
8. Montoliu, R., Pla, F., Klaren, A.: Illumination Intensity, Object Geometry and Highlights Invariance in Multispectral Imaging. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3522, pp. 36–43. Springer, Heidelberg (2005)
9. Negahdaripour, S.: Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(9), 961–979 (1998)
10. Odobez, J.M., Bouthemy, P.: Robust multiresolution estimation of parametric motion models. Int. J. Visual Communication and Image Representation 6(4), 348–365 (1995)
11. Szeliski, R., Coughlan, J.: Spline-based image registration. International Journal of Computer Vision 22(3), 199–218 (1997)
12. Torr, P.H.S., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding 78, 138–156 (2000)