The Lightspace Change Constraint Equation (LCCE) with practical application to estimation of the projectivity+gain transformation between multiple pictures of the same subject matter

S. Mann, C. Manders, and J. Fung
University of Toronto, Dept. Electrical and Computer Engineering,
Room 2001, 10 King's College Road, Mailstop SFB540, M5S 3G4,
(416) 946-3387; mann@eecg.toronto.edu

Abstract

For many years the Brightness Constancy Constraint Equation (BCCE) has been used for optical flow and related computer vision computations. However, almost all cameras have some kind of automatic exposure feature such as Automatic Gain Control (AGC), so that the overall exposure level of the image varies as the camera is aimed at brighter or darker portions of a scene. Moreover, because most cameras have some kind of unknown nonlinear response function, the change due to AGC cannot be captured by merely applying a multiplicative constant to the pixels of each image. We propose, therefore, a Lightspace Change Constraint Equation (LCCE) that accounts for exposure change (AGC) together with the nonlinear response function of the camera. The response function can be automatically "learned" by an intelligent image processing system presented with differently exposed capture of the same subject matter in overlapping regions of registered images. Most importantly, a Logarithmic Lightspace Change Constraint Equation (LLCCE) is shown to have a very simple mathematical formulation. The LCCE (and Log LCCE) is applied to the estimation of the projective coordinate transformation between pairs of images in a sequence, and is compared with examples where the BCCE fails.

1 Computer vision in lightspace (quantimetric imaging)

Commonly, calculations of motion estimation performed on images taken of the same subject with differing spatial alignment are done using pixels. In past work, Mann has shown that pixels are inadequate for many such calculations[2]. Rather, it is argued that photoquantites are better suited to such calculations. In this paper we use photoquantites for motion estimation to yield better results than motion estimation using pixels.

Consider Fig. 1 which illustrates how a camera with automatic exposure control takes in a typical scene. As we look straight ahead we see mostly sky, and the exposure is quite small. Looking to the right, at darker subject matter, the exposure is automatically increased. In general, automatic exposure cameras tend to capture the same subject matter at different exposures, depending on how the pictures are framed. Since the differently exposed pictures depict overlapping subject matter, we have (once the images are registered in regions of overlap) differently exposed pictures of identical subject matter. In the example illustrated in Fig 1, we have six differently exposed pictures depicting parts of the University College building and surroundings.

Case	**Scene**	**Camera**
1:	3-D	zoom, rot., pan, tilt, fixed COP
2:	planar	zoom, rot., pan, tilt, free to translate

Table 1: Fixed Center Of Projection (COP) and rigid planar patch: The projective model is exactly justified for two "no parallax" cases for a static scene. The second case is useful in tracking, and possibly filtering out rigid planar patches such as advertisments that occur on billboards, as illustrated in Fig 4.

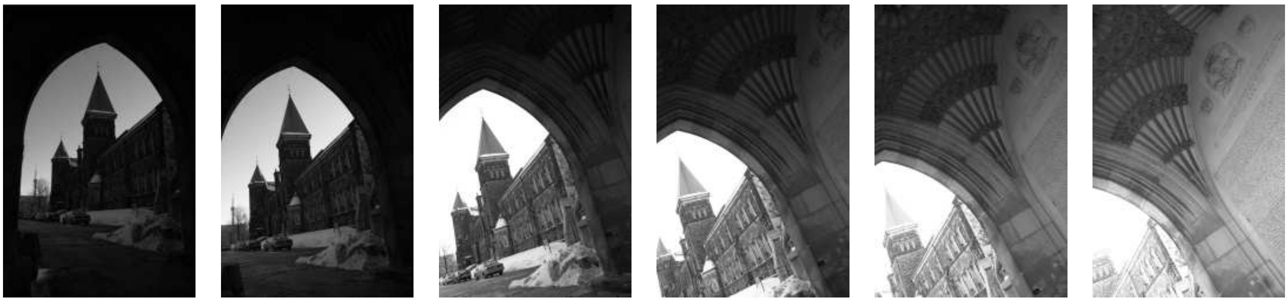
In this figure, the relationship between these pictures is assumed to be given by:

$$p_i = f\left(k_i q\left(\frac{\mathbf{A_i x} + \mathbf{b_i}}{\mathbf{c_i x} + d_i}\right)\right), \tag{1}$$

where f is the camera's response function[2], $\mathbf{x} = (x, y)$ denotes the spatial coordinates of the image, k_i is a single unknown scalar exposure constant, and parameters $\mathbf{A}_i$, $\mathbf{b}_i$, $\mathbf{c}_i$, and d_i denote the projective coordinate transformation between successive pairs of images: $\mathbf{A} \in \mathbf{R}^{2\times 2}$ is the linear coordinate transformation (e.g. accounts for magnification in each of the x and y directions and shear in each of the x and y directions), b is the translation in each of these two coordinate directions, and c is the projective chirp rate in each of these two coordinate directions[1]. This model is justified exactly under two specific cases (Table 1) where the second case, pertaining to rigid planar patches, is used, for example, for processing portions of images containing a billboard or other flat surface, an example of which is shown in Fig 4. The additional constant d makes the coordinate transformation into a group[12].

Without loss of generality, k_0 will be called the reference exposure, and will be set to unity, and frame zero will be called the reference frame, so that $p_0 = f(q)$. Thus we have:

$$\frac{1}{k_i} f^{-1}\left(p_i\left(\frac{\mathbf{A_i x} + \mathbf{b_i}}{\mathbf{c_i x} + d_i}\right)\right) = f^{-1}(p_0), \forall i, 0 < i < I. \tag{2}$$



(a) (b) (c) (d) (e) (f)

Figure 1: Automatic exposure as the cause of differently exposed pictures of the same (overlapping) subject matter. (a) Looking from inside Hart House Soldier's Tower, out through an open doorway, when the sky is dominant in the picture, the exposure is automatically reduced, and we can see the texture (clouds, etc.) in the sky. We can also see University College and the CN Tower, off in the distance, to the left. (b-e) As we look up and to the right, to take in subject matter not so well illuminated, the exposure automatically increases somewhat. We can no longer see detail in the sky, but new architectural details inside the doorway start to become visible. (f) As we look further up and to the right, the dimly lit interior dominates the scene, and the exposure is automatically increased dramatically. We can no longer see any detail in the sky, and even the University College building, outside, is washed out (over-exposed). However, the inscriptions on the wall (names of soldiers killed in the war) now become visible.

Taking the logarithm of both sides,

$$F^{-1}\left(p_i\left(\frac{\mathbf{A_i x} + \mathbf{b_i}}{\mathbf{c_i x} + d_i}\right)\right) - K_i = F^{-1}(p_0), \forall i, 0 < i < I, \tag{3}$$

where $K = log(k)$, and F^{-1} is the logarithmic inverse camera response function.

Re-arranging, we have:

$$F^{-1}\left(p_i\left(\frac{\mathbf{A_i x} + \mathbf{b_i}}{\mathbf{c_i x} + d_i}\right)\right) - F^{-1}(p_0) = K_i, \forall i, 0 < i < I. \tag{4}$$

This relation suggests a way to estimate the camera response function, f, from a pair of differently exposed images of overlapping subject matter, once the images are spatially registered.

Photographic film is traditionally characterized by the so-called "Density versus log Exposure" *characteristic curve* [3][4]. Similarly, in the case of electronic imaging, we may also use logarithmic exposure units, $Q = \log(q)$, so that, in areas of overlapping but identical subject matter, once the images are registered, one image will be $K = log(k)$ units darker than the other:

$$\log(f^{-1}(p_1(\mathbf{x}))) = Q = \log\left(f^{-1}\left(p_2\left(\frac{\mathbf{Ax} + \mathbf{b}}{\mathbf{c^T x} + d}\mathbf{x}\right)\right)\right) - K \tag{5}$$

The existence of an inverse for f follows from a semi-monotonicity assumption [5][6][7][2]. The goal is to estimate $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$, K, and F, from various input images, which can be achieved using a generalization of motion estimation, within an iterative framework, as follows:

- First compute the comparagrams between successive pairs of images. The comparagram is a two dimensional array c_{ij} that counts how many times a pixel in the first image has a value i at the same spatial location that a pixel in the second image has the value j. Comparagrams capture everything that can be known about the amplitude response function f of a camera [5][6][7][8];
- Slenderize the comparagrams to obtain the comparametric functions (comparagraphs) [5][6][7];
- Unroll the comparagraphs to obtain an estimate of the response function. The unrolling can be done using the GNU Unrolling utility from the comparametrics toolkit (available online, from comparametric.sourceforge.net);
- Use this estimate of the response function to determine the photographic quantities q_i (i.e. arrays of lixel values);
- Optional step: Tonally align the images (e.g. adjust the effective exposures). This tonal alignment need not be too accurate as this is simply a first guess (this step need not even be done);
- Estimate the spatiotonal transformation between successive pairs of images;
- Use the now registered images to obtain better estimates of the comparagrams. Repeat as often as necessary.

2 Motion estimation with the the Lightspace Change Constraint Equation (LCCE)

Hundreds of papers have been published on the problems of motion estimation and frame alignment [9], and much of this work is based on the so-called Brightness Constancy Constraint Equation (BCCE) of Horn and Schunk [10]. However, a more general formulation, suitable for mediated reality, is based on the Lightspace Change Constraint Equation (LCCE) [2].

Tsai and Huang [11] pointed out that the elements of the projective *group* give the true camera motions with respect to a planar surface. Tsai and Huang explored the group structure associated with images of a 3-D rigid planar patch, as well as the associated *Lie algebra*, although they assume that the correspondence problem has been solved. The solution presented in this paper (which does not require prior solution of correspondence) also relies on projective group theory [12].

When the change from one image to another is small, optical flow [10] may be used. In 1-D, the traditional optical flow formulation assumes each point x in frame t is a translated version of the corresponding point in frame $t + \Delta t$, and that Δx and Δt are chosen in the ratio $\Delta x / \Delta t = u$, the translational flow velocity of the point in question.

It be shown in the computations below, that it is advantageous to use the log quantimetric quantities, $Q(x)$ at time t, which is $\mathbf{Q}(x,t)$ in calculations of quantimetric flow [2] on $\mathbf{Q}(x,t)$ as described by:

$$\mathbf{Q}(x,t) = \mathbf{Q}(x + \Delta x, t + \Delta t) + K, \ \ \forall (x,t), \qquad (6)$$

where u is the translational flow velocity, and $\mathbf{x} = [x, y]$ is the spatial coordinate, and $\mathbf{Q}(x,t)$ is the logarithm of the camera response function of the log photoquantity, $log(q)$. This of course modifies and generalizes the Brightness Change Constraint Equation, because we could obviously evaluate the camera response function F at both sides of (6):

$$F(\mathbf{Q}(x,t)) = F(\mathbf{Q}(x + \Delta x, t + \Delta t) + K). \qquad (7)$$

However, for simplicity, it will now be understood that we will work in lightspace Q rather than imagespace F.

Expanding the right hand side of (6) in a Taylor series, and canceling 0th order terms gives: $u\mathbf{Q}_x + \mathbf{Q}_t + h.o.t. = 0$, where $\mathbf{Q}_x = d\mathbf{Q}(x,t)/dx$ and $\mathbf{Q}_t = d\mathbf{Q}(x,t)/dt$ are the spatial and temporal derivatives respectively, and $h.o.t.$ denotes higher order terms. Typically, the higher order terms are neglected, giving the expression for the lightflow at each point in one of the two images:

$$u\mathbf{Q}_x + \mathbf{Q}_t \approx 0 \qquad (8)$$

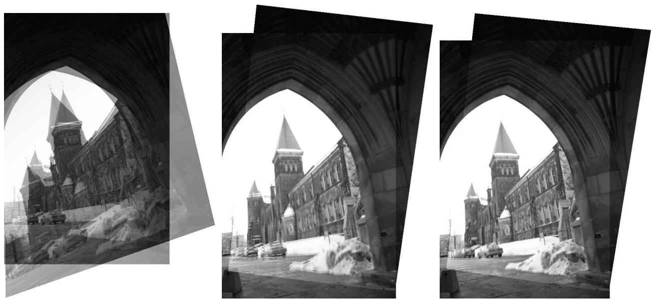
However, when automatic gain control is involved (e.g. when there can be a gain change between successive frames of video), we have

$$u\mathbf{Q}_x + \mathbf{Q}_t \approx -K \qquad (9)$$

For simplicity in the mathematical notation, we illustrate the situation for one dimensional "images", using the projective coordinate transformation $(ax + b)/(cx + d)$, whereas we recognize that with $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{c}$ there are actually eight scalar parameters in place of the three a, b, and c. Neglecting the special case in which the camera turns a full 90 degrees between two successive frames of video, we set $d = 1$.

Consider the lightflow velocity given by (9) For 'projective-flow' ('p-flow') [2], substitute $u = \frac{ax+b}{cx+1} - x$ into the BCCE (9). We may estimate the parameters of projectivity in a simple manner, based on solving a linear system of equations. To do this, we write the Taylor series of u:

$$u + x = b + (a - bc)x + (bc - a)cx^2 + (a - bc)c^2x^3 + \cdots \qquad (10)$$



(a) (b) (c)

Figure 3: Comparison of the proposed method with traditional methods. (a) Motion estimation in imagespace, e.g. by using the Brightness Constancy Constraint Equation (BCCE) for motion estimation between p_1 and p_2 gives poor results when there is a large jump from one image to the next. (b) Parameter estimation in lightspace, e.g. by using the Lightspace Change Constraint Equation (LCCE) to estimate the motion between $f^{-1}(p_1)$ and $f^{-1}(p_2)$ gives much better results. (c) Parameter estimation in log lightspace, e.g. by using the Log Lightspace Change Constraint Equation (LLCCE) to estimate the motion between $F^{-1}(p_1)$ and $F^{-1}(p_2)$ gives the best results. Moreover, the mathematical formulation in log lightspace is much simpler than that of either imagespace or linear lightspace.

and use the first 3 terms, obtaining enough degrees of freedom to account for the 3 parameters being estimated. Letting the squared error due to higher order terms in the Taylor series approximation be $\varepsilon = \sum(-h.o.t.)^2 = \sum((b + (a - bc - 1)x + (bc - a)cx^2)\mathbf{Q}_x + \mathbf{Q}_t)^2$, $\mathbf{q}_2 = (bc - a)c$, $\mathbf{q}_1 = a - bc - 1$, and $\mathbf{q}_0 = b$, and differentiating with respect to each of the 4 parameters of $\mathbf{q}$, setting the derivatives equal to zero, and solving, gives the linear system of equations for projective flow with gain:

$$\begin{bmatrix} \sum x^4\mathbf{Q}_x^2 & \sum x^3\mathbf{Q}_x^2 & \sum x^2\mathbf{Q}_x^2 & \sum x^2\mathbf{Q}_x \\ \sum x^3\mathbf{Q}_x^2 & \sum x^2\mathbf{Q}_x^2 & \sum x\mathbf{Q}_x^2 & \sum x\mathbf{Q}_x \\ \sum x^2\mathbf{Q}_x^2 & \sum x\mathbf{Q}_x^2 & \sum \mathbf{Q}_x^2 & \sum \mathbf{Q}_x \\ \sum x^2\mathbf{Q}_x & \sum x\mathbf{Q}_x & \sum \mathbf{Q}_x & \sum 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ K \end{bmatrix} = - \begin{bmatrix} \sum x^2\mathbf{Q}_x\mathbf{Q}_t \\ \sum x\mathbf{Q}_x\mathbf{Q}_t \\ \sum \mathbf{Q}_x\mathbf{Q}_t \\ \sum \mathbf{Q}_t \end{bmatrix} \qquad (11)$$

where $\alpha = b$, $\beta = (a - bc)$, and $\gamma = (bc - a)c$, and where the two dimensionality of the images is, for simplicity, not shown explicitly.

Note that this set of equations captures both domain and range motion simultaneously.

2.1 Composing Images with Certainty Functions

The inverse camera response function f^{-1} can be used to convert pixel values into photographic light quantities (lixel values), and vice-vera (with f).

Once the camera response function is estimated, along with the projective coordinate transformations between successive frames of video, the images may be brought together into a common coordinate space, as shown in Fig 2.

This is done by pairwise estimation of the spatiotonal coordinate transformation between successive pairs of images in the video sequence.

For an example comparison between existing methodology, namely the use of the BCCE, and the new methods (LCCE and LLCCE) see Fig. 3.

3 Conclusions

A new Log Quantimetric image motion has been shown to be as easy to compute as traditional optical flow, yet deliver

Figure 2: Successive video frames of Fig. 1 are spatiotonally aligned, so that they all appear in the same spatial (projective) coordinates, as well as in the same tonal range (e.g. the same k value).

(a) (b) (c) (d)

Figure 4: (a)(b)Two frames from a video sequence in New York city, showing how a nice view of the Empire State building is spoiled by an offensive jeans wear advertisement (e.g. a billboard depicting a man pulling off a women's clothes). Notice how the effect of AGC is similar to that depicted in Fig. 1. (a) Since a large proportion of sky is included in the image, the overall exposure is quite low, so the image is darker. (b) Since the darker billboard begins to enter the center portion of view, the gain increases and the entire image is lighter. (c)(d)Two frames from the video sequence after it has been modified (filtered). Subject matter along the planar patch of the billboard is replaced with a view of digital eyeglasses (such as might be worn to filter such ads). Notice how the exposure of the new matter, introduced into the visual field of view, tracks the exposure of the offensive advertising material originally present. (c) A large proportion of sky is included in the image, so the overall exposure is quite low, making the image darker. The additional material inserted into the image is thus automatically made darker, comparametrically, to match the scene it was written into. (d) Since the original image was lighter, in this frame, the new matter introduced into the visual reality stream is also made lighter, comparametrically, to match.

much better results.

References

[1] S. Mann, "Joint parameter estimation in both domain and range of functions int the same orbit of the projective-wyckoff group," *IEEE International Conference on Image Processing (ICIP–96)*, pp. 193–196, December 1996. Also appears in: M.I.T. M.L. T.R. 384, 1994.

[2] S. Mann, *Intelligent Image Processing*. John Wiley and Sons, November 2 2001. ISBN: 0-471-40637-6.

[3] C. W. Wyckoff, "An experimental extended response film," *S.P.I.E. NEWSLETTER*, pp. 16–20, JUNE-JULY 1962.

[4] C. W. Wyckoff, "An experimental extended response film," Tech. Rep. NO. B-321, Edgerton, Germeshausen & Grier, Inc., Boston, Massachusetts, MARCH 1961.

[5] F. M. Candocia, "A least squares approach for the joint domain and range registration of images," *IEEE ICASSP*, vol. IV, pp. 3237–3240, May 13-17 2002. avail. at http://iul.eng.fiu.edu/candocia/Publications/Publicati ons.htm.

[6] F. M. Candocia, "Synthesizing a panoramic scene with a common exposure via the simultaneous registration of images," *FCRAR*, May 23-24 2002. avail. at http://iul.eng.fiu.edu/candocia/Publications/Publicati ons.htm.

[7] A. Barros and F. M. Candocia, "Image registration in range using a constrained piecewise linear model," *IEEE ICASSP*, vol. IV, pp. 3345–3348, May 13-17 2002. avail. at http://iul.eng.fiu.edu/candocia/Publications/Publicati ons.htm.

[8] S. Mann, "Compositing multiple pictures of the same scene," in *Proceedings of the 46th Annual IS&T Conference*, (Cambridge, Massachusetts), pp. 50–52, The Society of Imaging Science and Technology, May 9-14 1993. ISBN: 0-89208-171-6.

[9] S. B. J.L. Barron, D.J. Fleet, "Systems and experiment performance of optical flow techniques," *International journal of computer vision*, vol. 12, no. 1, pp. 43–77, 1994.

[10] B. Horn and B. Schunk, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[11] R. Y. Tsai and T. S. Huang, "Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch I," *IEEE Trans. Accoust., Speech, and Sig. Proc.*, vol. ASSP(29), pp. 1147–1152, December 1981.

[12] M. Artin, *Algebra*. Prentice Hall, 1991.