LIP Model Framework Applied to Fast Registration of Images Acquired from a Moving Camera under Variable Lighting

D'Arras, Paul[(*)(**)]*,Becker, Jean-Marie*[(*)]*, Jourlin, Michel*[(*)] *and Bouabdellah, Mohamed*[(**)]

[(*)]*Laboratoire Hubert Curien, St Etienne University, France*
[(**)]*NT2I Company, St Etienne, France*

p.darras@nt2i.fr

Keywords

Fast Registration, variable lighting, LIP model, moving camera, ground vehicle.

Introduction

We get a video stream from an embedded camera on a moving ground vehicle, in side view and *with strong lighting variations*. In order to perform a panoramic visualization of this stream, we focus on the calculation of the translation vector between two successive images. To address the issue of lighting variations, we have used operators defined in the LIP (Logarithmic Image Processing) framework. The produced algorithms are expected to be robust, simple and highly parallelizable, with the ultimate goal (not covered in this paper) to implement these algorithms in VHDL on a FPGA configuration.

Materials and Methods

The frame rate is assumed sufficiently high to permit a local assimilation of the movement to a simple translation. In order to get enough information in the common area of two consecutive images, we impose an overlap area of at least 30% of the image size (Figure 1).

Color images are dealt with a separate processing for each channel.

<u>Two consecutive side views with lighting change (Figure 1)</u>
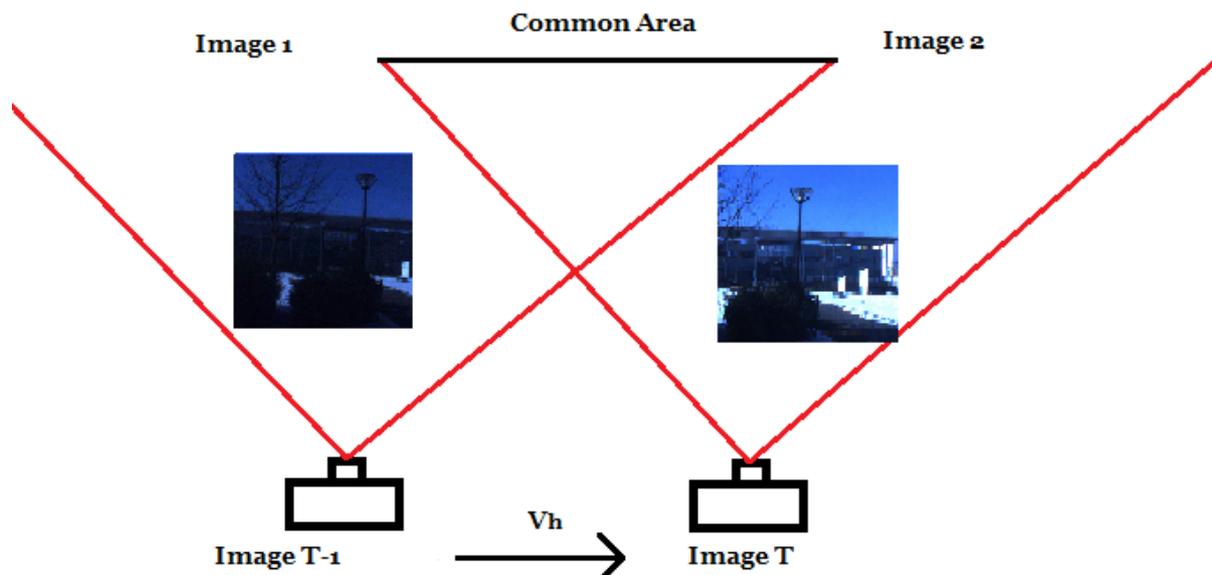
Determination of the translation vector for two successive images

If I_1 and I_2 have the same size $W*H$, we determine the following L1-distances:

$$D(k,l) = \sum_{i,j}(|I_1(i,j)-I_2(i+k,j+l)|),\ 0\leq i+k<W, \qquad 0\leq j+l<H, \qquad 0\leq i<W, \qquad 0\leq j<H.\ (Eq:1)$$

Université de Liège

The desired translation vector $V=[v_x, v_y]$ is given by the minimum of $D(k,l)$:):

$$min_{(k,l)}(D(k,l)) = D(k_{min}, l_{min}).$$

The translation vector is thus $V = [k_{min}, l_{min}]$.

This method remains extremely slow even when parallelized.

In order to speed up the computation of V, an attempt was made to obtain the minimum $D(k,l)$ distance on points sampled every 2^k pixels. This procedure is repeated in a dichotomous manner, on points sampled 2^{k-1} pixels around the minimum found on the previous iteration, and then 2^{k-2} pixels sampling… Unfortunately, this method, prone to be caught in local minima, must be rejected.

Finally a successful approach was found to be a pyramidal technique with a quad-tree data structure [1], beginning by a rough evaluation of the translation vector $V_o = [vx_o, vy_o]$ on sub-sampled images, this evaluation being iteratively refined on windows with increasing resolution.

Light stabilization

The determination of the translation vector can be strongly impaired by variable lighting between two neighbor frames. A pre-processing is done on each image in order to bring the two successive images under a comparable lighting.

This light stabilization is based on the LIP (Logarithmic Image Processing) [2] model operators $\triangle\!\!\!\!+$ and $\triangle$ (cf. Annex). In fact it has been established in [3] that the logarithmic addition or subtraction of a constant simulate lighting variation. The main steps are as follows.

A target mean $Mref$ is chosen (for example: grey level 128) for the stabilization.

Let I be the original image, $M(I)$ its mean and Is the stabilized image with $M(Is) = Mref$.

If $M(I) < Mref$, let $C = \dfrac{Mref - M(I)}{1 - \frac{M}{256}}$ then $Is = I \mathbin{\triangle\!\!\!\!+} C$.

If $M(I) > Mref$, let $C = \dfrac{M(I) - Mref}{1 - \frac{M(I)}{256}}$ then $Is = I \mathbin{\triangle} C$.

Distance homogenization

At this step, each image is globally stabilized, but two successive images may not be stabilized in the same way; this is the case when one of them has a saturated area without any similar zone in the other. This drawback, due to the global stabilization, can be corrected by replacing the usual image subtraction in (Eq 1) should be replaced by a LIP subtraction [4]:

$$D(k,l) = \sum_{i,j}(|I1(i,j) \mathbin{\triangle} I2(i+k,j+l)|), \qquad 0 \le i+k<W, \qquad 0 \le j+l<H, \qquad 0 \le i<W, \qquad 0 \le j<H.$$

Results and Discussion

The algorithms have been implemented in C++ on a computer with the following characteristics: Intel core i7-3610QM, CPU @2.30GHz and Windows 64 bits with an 8 Gbytes RAM.

The test images size is 958*958 px. An example of registration of two images in presence of strong lighting variation is given in the Figure 2.

Paul D'Arras

Figure 2: Examples

(a) image 1 of figure 1 (Note that this image is obtained by closing the lens iris), (b) image 2 of figure 1, (c) image 1 stabilized, (d) image 2 stabilized, (e) panorama reconstruction. One can notice the slight vertical component of the translation vector (vertical arrow).
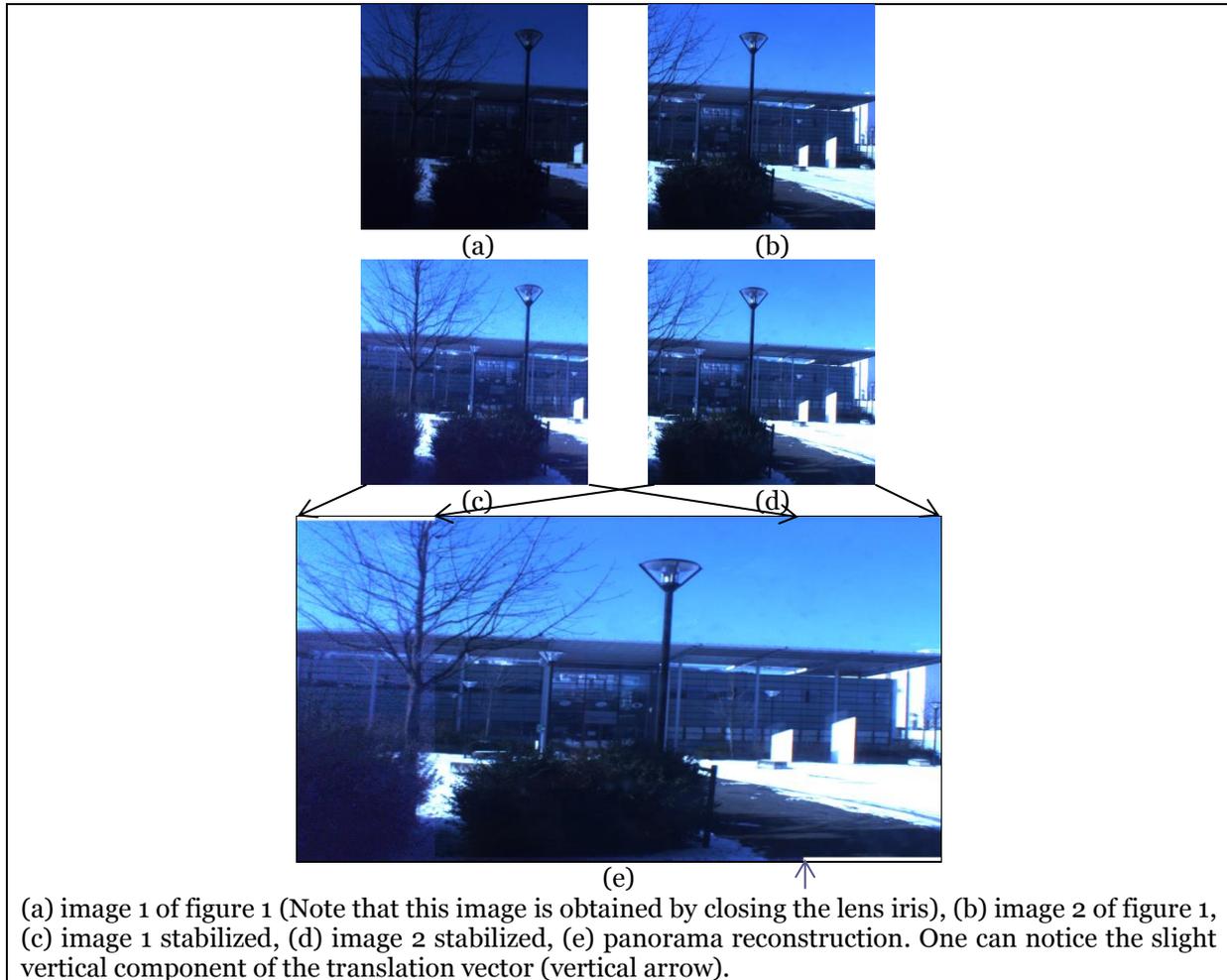
Conclusion and perspectives

Our method combines a pyramidal algorithm, resulting in a good computational time, and the LIP model, allowing an efficient registration even with sudden lighting changes.

The target of the registration being a human operator, the objective of a visually satisfying reconstruction is reached [5] due to the consistency of the LIP model with human vision.

In this presentation, no a priori knowledge has been taken into account, though a certain amount of information can be exploited from one frame to the other. The common ROI of the two images can be predicted by the vehicle speed and the camera frame rate; in the same way the translation vector is predictable as long as the scene depth varies slowly.

If the vehicle's road is very uneven, sudden shocks may appear. In this case, the registration can no longer be reduced to a simple translation, implying a more complex kinetic model.

Our ultimate goal will be to reach the real time barrier by developing all our algorithms in VHDL in order to benefit from their parallelism and the computing power of FPGA boards.

Annex

The LIP Model is now recognized as an efficient framework possessing strong physical and mathematical properties. Based on the transmittance law, it is perfectly adapted to process images acquired in transmission. Moreover, thanks to its consistency with Human Vision, it also permits to interpret images as a reflected image like a human eye would do.

The two basic logarithmic operations of the LIP model are the addition of two images f and g, denoted f $\triangle\kern-0.9em+$ g and the multiplication of an image f by a real number λ, noted λ $\triangle\kern-0.9em\times$ f:

$$f \triangle\kern-0.9em+\; g = f + g - \frac{f*g}{M}$$ where M represents the grey scale size (here 256)

$$\lambda \triangle\kern-0.9em\times\; f = M - M\left(1 - \frac{f}{M}\right)^{\lambda}$$

Such operators give the space of images a Vector Space structure, which permits to define logarithmic interpolation, new notions of contrasts and metrics, scalar product of two images...

The subtraction of two images f and g, where f ≥ g is given by:

$$f \triangle\kern-0.9em-\; g = \frac{f - g}{1 - \frac{g}{M}}$$

References

[1] Tonye E., Akono A., Jolion J.M., (1994) 'Approche objet et pyramidale dans la classification non supervisée des images en télédétection', Tankoano J. (ed.), Proceedings of the second African conference on research in computer science, pp.203-217, Ouagadougou.

[2] Jourlin M. and Pinoli J.C. (1988) 'A model for logarithmic image processing' in Journal of Microscopy, V.149, January 1988, pp.21-35.

[3] Carré M. and Jourlin M. (2014) 'LIP operators: simulating exposure variations to perform algorithms independent of lighting conditions'. International Conference, Multimedia Computing and Systems (ICMCS), Marrakech.

[4] Jourlin M., Carré M., Breugnot J., Bouabdellah M., (2012), 'Logarithmic Image Processing: Additive Contrast, Multiplicative Contrast, and Associated Metrics', Advances in Imaging and Electron Physics, Vol. 171, pp. 357-406.

[5] Brailean J.C. (1991) 'Evaluating the EM algorithm for image processing using a human visual fidelity criterion'. International Conference on Acoustics, Speech, and Signal Processing, pp 2957-2960, Toronto, Ont.