Computación y Sistemas Vol. 6 No. 4 pp. 293 - 299
© 2003, CIC - IPN. *ISSN 1405-5546* Impreso en México

Qualitative Interpretation of Camera Motion in the Projection of Individual Frames of an Image Stream

Interpretacion Cualitativa del Movimiento de la Camara mediante el Analisis de Projecciones de los Cuadros de una Secuencia de Imagenes

Joaquín Salas

CICATA-IPN
Jose Siurob 10, Col Alameda
Queretaro, Qro. CP 76040
E-mail: salas@ieee.org

Article received on March 19, 2002; accepted on March 03, 2003

Abstract

In this study, we are interested on some special kind of camera movements that ideally take place on a flat surface and the image projections perpendicular to these movements. We analyze interpretation of motion for a camera heading forward, panning around its optical center and translating perpendicular to its optical axis. We present simulations and experiments with real data. Our results show that it is possible to obtain qualitative information about the camera motion's nature.

Keywords: Camera Motion, Qualitative Interpretation, Image Sequence, Flatland

Resumen

En este documento, estudiamos una clase especial de movimientos de la cámara que idealmente se desarrollan en una superficie plana y las proyecciones de imagen perpendiculares a esos movimientos. De esta forma, analizamos la interpretación de movimientos de una cámara cuando ésta avanza en dirección de su eje focal, gira alrededor de su centro óptico y se traslada perpendicular a su eje focal. Mostramos simulaciones y experimentos con datos reales. Nuestros resultados muestran que la aproximación permite obtener información cualitativa sobre la naturaleza del movimiento de la cámara.

Keywords: Movimiento de la Cámara, Interpretación Cualitativa, Secuencias de Imágenes, Mundo Plano

1 Introduction

Computer Vision is concerned with the tridimensional interpretation a scene from a sequence of images. For one thing, this problem is important because in theory, there is an infinity number of objects that can produce the same image, *e.g.*, varying the object size and the distance from the camera to the object. In particular, we investigate some uses of projections of individual frames that are part themselves of an image stream. In a way similar to images, projections are not unique. Nonetheless, in this document, we explore some limits where this notion can be challenged in practice. Projections are well known compact representations of images[Jain et al., 1995](see Fig. 4(a) and its projection in Fig. 2(a)). In a compact image stream the variations of projections of individual frames are small. Thus providing almost unique characteristics to a particular camera trajectory.

In some cases, camera motion is solved along with scene structure. Nevertheless, computing structure from motion has been shown to be an extremely difficult problem. However, significatively advances have been done in the area. For instance, Tomasi[Tomasi, 1991] developed an optimal solution for the case of ortographic projection. Under perspective projection, extreme care must be given to computing the intrinsic and extrinsic camera parameters. Even then, the solution is brittle and numerically unstable. Today's state of the art includes making Euclidean reconstruction from basically uncalibrated cameras[Kahl and Heyden, 2001]. The two dominant approaches are factorization-like methods[Zhang and Tomasi, 1999] for weak-perspective and iterative solutions with Kalman filtering[Kim et al., 1997]. The former requires solving sequential matching while the latter involves the problem of establishing a good initial starting point for

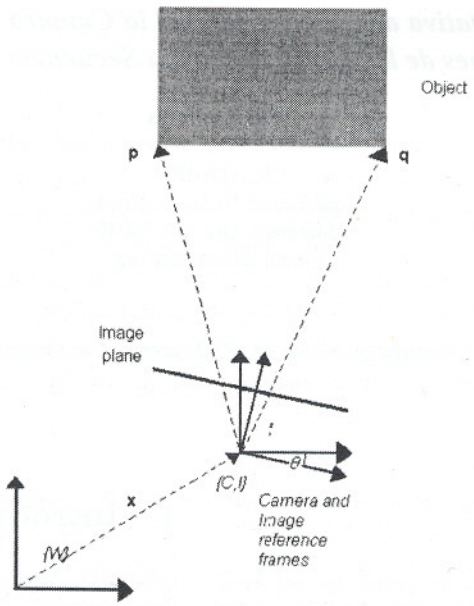


Figure 1: Imaging object points in flatland. An object with corners $\mathbf{p}$ and $\mathbf{q}$ is placed in the world. The world reference system is centered in $\{W\}$. A camera, with reference system $\{C\}$, has its optical center in $\mathbf{x}$. The camera focal length is f. The image's reference system is placed in $\{I\}$. The angle θ is the angular difference between $\{C\}$ and the image reference system $\{I\}$.We would like to verify how object points are imaged as the camera moves in its workplace.

bundle adjustment. Duric and Rivlin [Duric et al., 2000] analyze what happen with the histogram of normal optical flow when the camera rotates, translates in the direction of the optical axis, perpendicular to the optical axis and around the axis perpendicular to the optical axis. Since obtaining complete structure from motion has shown to be difficult and error prone, we claim that it is worth pursuing trying to gather at least partial and qualitative information about the nature of camera motion.

In §2, we study ideal projection of object points for different types of camera motion. Next in §3, we present an scheme to track features along the projected individual frames in an image stream. Then in §4, we present some experimental results with both ideal and real data. Finally, we conclude with some remarks and discussion about research directions.

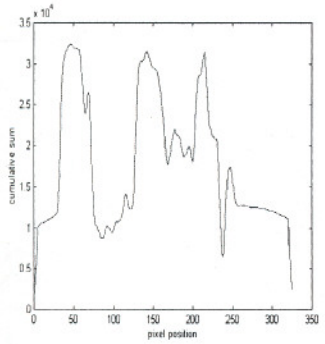
2 Imaging in Flatland

We are interested on a special kind of camera movements that ideally take place on a flat surface and the image projections perpendicular to these movements. Under these circumstances, the projection process may be described analytically by the Radon transform[Kak and Slaney, 1988], given by

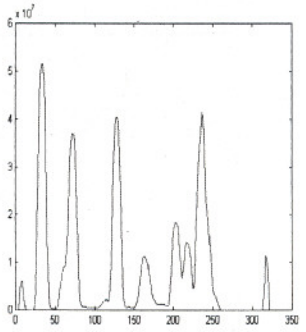
$$P_\gamma(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)\delta\left(x\cos\gamma + y\sin\gamma - t\right)dx\,dy \tag{1}$$

where $I(x,y)$ describes the image, $\delta(x,y)$ is a delta function, $x\cos\gamma + y\sin\gamma - t$ is a collection of parallel rays that forms an angle γ with the y-axis, and t is a distance along the projection. From now on, we will focus on the case where $\gamma = 0$. Let us consider the situation where a robot moves while a vision system grabs images with a very small timestamp difference between frames. In this section, we review how world points will be imaged in noise free flatland(see Fig. 1). Unless stated all coordinates are expressed in the world reference system $\{W\}$. Suppose that an object with corners $\mathbf{p}$ and $\mathbf{q}$ is placed in the world. A camera, with reference system $\{C\}$, has its optical center in $\mathbf{x}$. The camera focal length is f. Finally, the image's reference system is placed in $\{I\}$. We would like to verify how object points are imaged as the camera moves in its workplace. The line between points $\mathbf{p}$ and $\mathbf{x}$ is given by

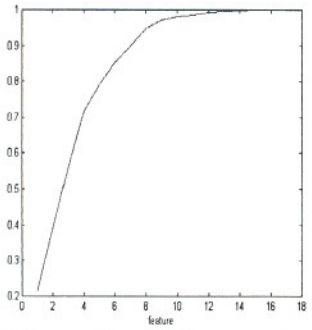
$$r\mathbf{p} + (1-r)\mathbf{x} = \mathbf{t} \tag{2}$$



(a) Projection of the image in Fig. 4(a)



(b)Result of evaluating Eq. (12) in plot (a)



(c) The value of the maximum in (b) are ordered from larger to smaller and divided by their cumulative sum.

Figure 2: Features are found in the projection of images. They are chosen among the larger values of Eq. (12). A threshold is established based on their cumulative sum.

where the points in the segment $\overline{\mathbf{px}}$ are the points $\mathbf{t}_0$ for which $r \in [0,1]$. On the other hand, the image plane is defined by the line

$$\rho = [\mathbf{s} + \mathbf{x}]^T \mathbf{u} \tag{3}$$

Here $\mathbf{u} = [\cos\theta, \sin\theta]^T$ is the vector in the direction of the shortest point from the image plane and both the camera $\{C\}$ and the world $\{W\}$ reference systems; ρ is the shortest distance from $\{W\}$ to the image plane; $\mathbf{s}$ is a point in the image plane; and θ is the angular difference from $\{C\}$ and the image reference system $\{I\}$. Both lines intersect when $\mathbf{s} + \mathbf{x} = \mathbf{t}$. That is when r equals

$$r = \frac{\rho - \mathbf{x}^T\mathbf{u}}{[\mathbf{p} - \mathbf{x}]^T\mathbf{u}} \tag{4}$$

This point is given in world coordinates. To get the projection in image coordinates, the following transformation applies

$$\mathbf{p}^I = T_C^I T_W^C \mathbf{p}^W \tag{5}$$

where T_B^A refers to a homogenous transform involving a rotation and a translation such that $\mathbf{p}^A = T_B^A \mathbf{p}^B = R_B^A \mathbf{p}^B + \mathbf{t}^A$. Therefore, given a known camera motion, we have a way to express a feature's projection. Otherwise stated, when the camera moves perpendicular to its optical axis, the projection $u_1(x)$ of an object point $\mathbf{p} = (x, z)$ is $u_1(x) = k_1 x$, where $k_1 = f/z$. When the camera moves along the direction of its optical axis is $u_2(z) = k_2/z$, where $k_2 = fx$. Finally, when pans around its optical center it is $u_3(\theta) = f\cot\theta$.

3 Motion Tracking

Shi and Tomasi[Shi and Tomasi, 1994] studied the problem of tracking two-dimensional image features from frame to frame using a Newton-Raphson type of search. In our case, the problem is simplified since we track one dimensional features. The following development is largely based on Shi and Tomasi for the case of one dimensional features. Let $J(x)$ and $I(x)$ be two consecutive frame projections. The dissimilarity between corresponding features separated a distance d can be measured by

$$\epsilon(d) = \int_F (J(x+d) - I(x))^2 dx \tag{6}$$

where F is a small interval over which similarity is sought. The term $J(x+d)$ can be expressed by the linear terms, neglecting the second and higher order terms, of its Taylor's expansion as

$$J(x+d) \approx J(x) + d\frac{\partial J(x)}{\partial x} \tag{7}$$

Thus,

$$\frac{\partial J(x+d)}{\partial d} \approx \frac{\partial J(x)}{\partial x} \tag{8}$$

The derivative of $\epsilon(d)$ is

$$\frac{\partial \epsilon(d)}{\partial d} = 2\int_F (J(x+d) - I(x))\frac{\partial J(x+d)}{\partial d} dx \tag{9}$$

Replacing $J(x+d)$ and its derivative by their approximation, we have

$$\frac{\partial \epsilon(d)}{\partial d} \approx 2\int_F \left(J(x) + \frac{\partial J(x)}{\partial x} d - I(x)\right)\frac{\partial J(x)}{\partial x} dx \tag{10}$$

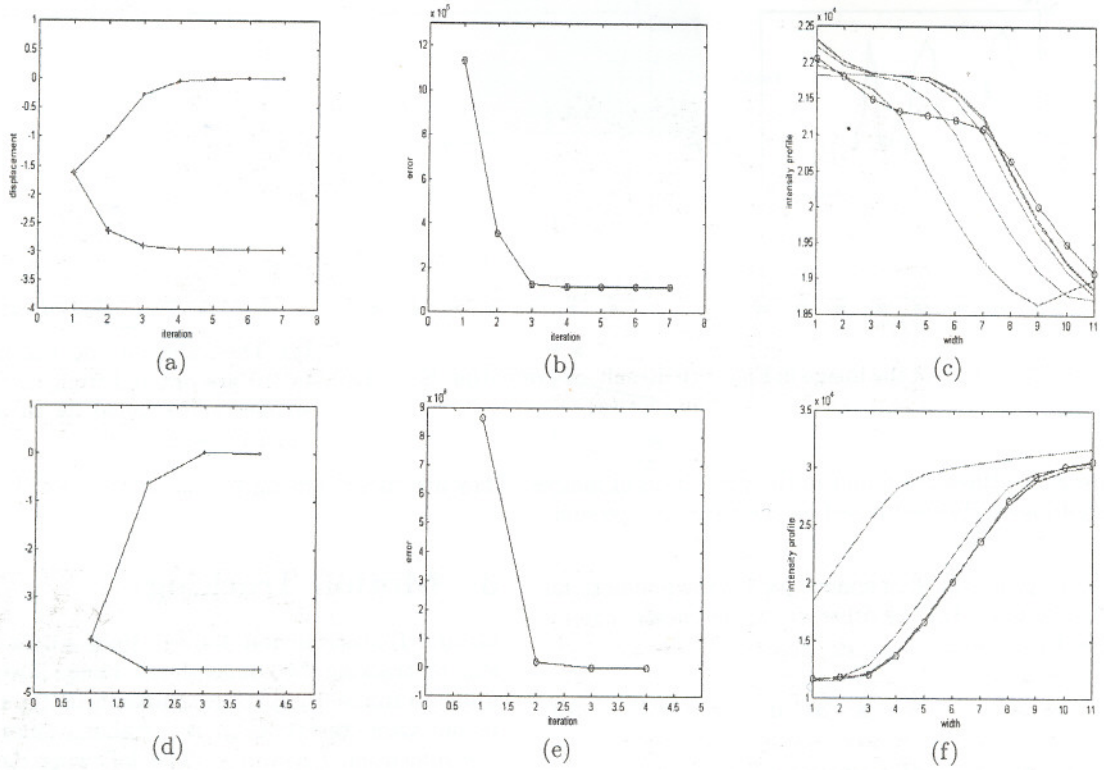


Figure 3: Matching image profile features. In 3(a), 3(b) and 3(c) we found the correspondence for a feature after 7 iterations. In 3(a) the upper line shows the displacement at each iteration. The line below it shows the accumulative displacement. The final displacement is -4.4875 pixels. In 3(b) we show how the error is decreasing. The final error is 1.123 millions. In 3(c) we show graphically how both curves converge iteration after iteration to the circled curve. In 3(d), 3(e) and 3(f) we found the correspondence for another feature after 4 iterations. In 3(d) the upper line shows the displacement at each iteration. The line below it shows the cumulative displacement. The final displacement is -2.9648 pixels. In 3(e) we show how the error is decreasing. The final error is 717,370. In 3(f) we show graphically how both curves converge iteration after iteration. The rightmost curve is the best fit to the circled curve.

The minimum error yields when the derivative of $\epsilon(d)$ equals zero. Therefore d can be expressed as

$$d = z^{-1}e \tag{11}$$

where

$$z = \int_F \left(\frac{\partial J(x)}{\partial x}\right)^2 dx \tag{12}$$

and

$$e = \int_F (J(x) - I(x)) \frac{\partial J(x)}{\partial x} dx \tag{13}$$

The value of z is a good reference about how easy it is to track a feature. That is, when its value is small the displacement is large and convergence may be poor. Contrariwise, when z is large, the iteration tends to converge.

Since non linear factors become important under most situations, Eq.(11) has to be replaced by the following iterative formulation

$$d_{k+1} = d_k + d \tag{14}$$

4 Experimental Results

The equations outlined in §2 can be used to get insight about the projection of object points in flatland under different types of camera motion. Suppose that there is an object with points $\mathbf{p} = [-4, 40]^T$ and $\mathbf{q} = [4, 40]^T$. The focal length is one unit. In Fig. 4 there are some resulting plots. In Fig. 4(b) the center of projection was moved between $[0, 3]^T$ and $[0, 30]^T$ units. The focal axis coincides with the direction of motion. The imaged object size is inversely proportional to the distance between

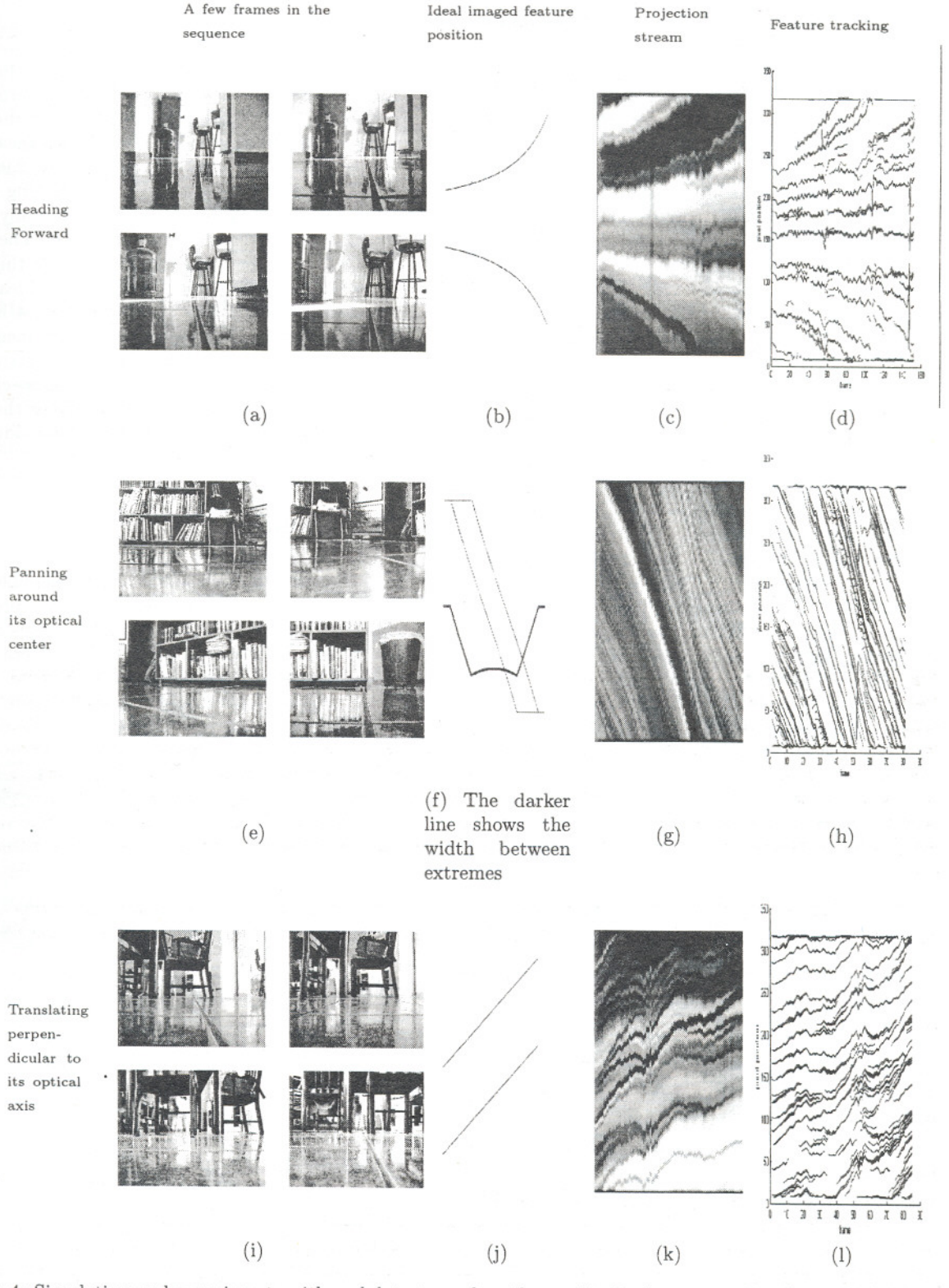


Figure 4: Simulation and experiments with real data to analyze the qualitative interpretation of motion for a camera heading forward, panning around its optical center and translating perpendicular to its optical axis.

the camera and the object. In Fig. 4(f) we rotated the camera between 50 and 130 degrees.

The center of projection is fixed at $\mathbf{x} = [0,3]^T$. In this range the variation is almost linear and may be confused with camera motion perpendicular to the optical axis. In 4(j) the camera moves laterally between $\mathbf{x}_i = [-3,0]^T$ and $\mathbf{x}_f = [3,0]^T$. The focal axis is perpendicular to the direction of motion. As reported by Bolles[Baker and Bolles, 1988], since the image size depends on the distance between the camera and the object and hence remains constant through the displacement under a predefined motion, the extended base line may provide both a way to a robust numerical solution and simple algorithm for tasks such as occluding boundary detection, stereo analysis and others.

Given the simulation results, we gather some experimental data. In Fig. 4, we show three dense sequences. Figure 4(c) corresponds to images of the scene shown in Fig. 4(a). It shows a sequence of 153 intensity profiles of a camera heading forward. Physically, the camera advanced about 120cms. There was not strict control on the distance between frames. Indeed the variations in orientations are remarkable. Nevertheless, the camera had an overall forward trajectory with its optical axis approximately in its direction of motion. Here, we observe how the features diverge as the camera gets closer. Figure 4(g) corresponds to images of the scene shown in Fig. 4(e). It presents a 82 intensity profiles when the camera rotates about 90 degrees. Again, the angle between frames is not equal. Also there is not warranty that the optical center coincides with the center of rotation. Finally, 4(k) corresponds to images of the scene shown in Fig. 4(i). It presents a dense sequence of 82 intensity profiles where we imaged in a direction perpendicular to the direction of motion. The camera advanced about 90cm.

The intensity profile of the image in Fig. 4(a) is shown in Fig. 2(a). Then in Fig. 2(b), we show the function z when the size of the window F is 10 pixels. The values of the maximum sorted by decreasing order as a percentage of the cumulative sum are given in Fig. 2(c). A good feature tends to be present when there is abrupt change in the profile intensity values. In the rest of the experiments, we consider a good feature to those which cumulative value are below 99% of the sum of the feature values.

Now, we may attempt to track a feature from line to line. In Fig. 3, we show a couple features of a given line and its tracking in the next line. In 3(a), 3(b) and 3(c) we found the correspondence after 7 iterations. In 3(a) the upper line shows the displacement at each iteration. The line below it shows the accumulative displacement. The final displacement is -4.4875 pixels. In 3(b) we show how the error is decreasing. The final error is 1.123 millions. In 3(c) we show graphically how both curves converge iteration after iteration to the circled line. In 3(d), 3(e) and 3(f) we found the correspondence for feature 1 after 4 iterations. In 3(d) the upper line shows the displacement at each iteration. The line below it shows the cumulative displacement. The final displacement is -2.9648 pixels. In 3(e) we show how the error is decreasing. The final error is 717,370. In 3(f) we show graphically how both curves converge iteration after iteration. The rightmost curve is the best fit to the circled curve. At this point, we are in the position to track all the selected features from one frame to the following. Figure 4(d), 4(h) and 4(l) show the result of tracking through the image stream. The lines are computed automatically by tracking the most promising features. These lines show clearly that it is possible to infer, at least qualitatively, the nature of camera motion from the projections of individual frames.

Conclusion

In this document, we show that it is possible to qualitatively interpret camera motion from the projection of individual frames in an image stream. This interpretation is made for the cases where the camera is moving in the direction of its optical axis, around its optical center, and perpendicular to its optical axis. Given an image streams with these type of camera motion, we presented a tracking scheme to follow features along the image sequence. It is possible to observe clearly how the difference from a rotation and a translation perpendicular to the direction of motion are generate very similar imaging features. The adds to the common believe that structure from motion is very sensitive in nature.

In this study, we show that useful information can be processed efficiently due to the compact representation of images and the smooth variation of the cumulative sum between frames. Further work aim to organize the redundant visual perception, and to quantitatively interpret camera motion, and to use this information to localize the camera in the workspace to allow visual based navigation.

Acknowledges

The author wants to thank to Prof. Carlo Tomasi for all his ideas, help and support and to the reviewers for their comments and suggestions. This work was partially supported with a grant from CEGEPI-IPN.

References

Baker, H. and Bolles, R. (1988). Generalizing Epipolar-Plane Image Analysis on the SpatiotemporalSurface. In *Computer Vision and Pattern Recognition*, pages 2-9.

Duric, Z., Rivlin, E., and Rosenfeld, A. (2000). Qualitative Description of Camera Motion from Histograms of Normal Flow. In *IEEE International Conference on Pattern Recognition*, volume~3, pages 194-198.

Jain, R., Kasturi, R., and Schunck, B. (1995). *Machine Vision*, McGraw Hill.

Kahl, F. and Heyden, A. (2001). Euclidean Reconstruction and Auto-Calibration from Continuous Motion. In *International Conference on Computer Vision*, volume 2, pages 572—577.

Kak, A.~C. and Slaney, M. (1988). *Principles of Computerized Tomographic Imaging*. IEEE Press.

Kim, E.T., Han, J.-K., and Kim, H.M. (1997). A Kalman-Filtering Method for 3-D Camera Motion Estimation from Image Sequences. In *IEEE International Conference on Image Processing*, volume 3, pages 630-633.

Shi, J. and Tomasi, C. (1994). Good Features to Track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593-600.

Tomasi, C. (1991). *Shape and Motion from Image Streams: a Factorization Method*. PhD thesis, Carnegie Mellon University. CMU-CS-91-172.

Zhang, T. and Tomasi, C. (1999). Robust and Consistent Camera Motion Estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 164-170.

***Joaquín Salas**, obteined his Doctor degree in Informatics from ITESM campus Monterrey in 1996. From then on, he has been affiliated with CICATA-IPN. He has been visiting scholar or invited profesor at Xerox PARC, Oregon State University, Universidad Autónoma de Barcelona, Stanford University, and the Ecole Nationale Superiore des Telecommunications. He has published 17 articles in international journals and congress on the general topic of image analysis. Since 1995, he has been member of Mexico's National System of Researchers. He was founding President of IEEE Querétaro, section and chairman of the IEEE 7th Mexico's National Minirobotics Contest*