REVIEW OF TECHNIQUES FOR MOTION ESTIMATION IN ARTIFICIAL VISION

REVISION DE TECNICAS PARA ESTIMACION DE MOVIMIENTO EN VISION ARTIFICIAL

Sandra Nope, Humberto Loaiza, Eduardo Caicedo

Universidad del Valle, Cali-Colombia
Ciudad Universitaria Meléndez, Calle 13 No 100-00
E-mail: {sandrano, hloaiza, ecaicedo}@univalle.edu.co

Abstract: Motion estimation has multiple applications in artificial vision from video compression and codification to perception tasks like surveillance of activities. We resume and classify the developed techniques for motion estimation. You can find a short description about the techniques, classification and some current works. There is a discussion in the end, where the advantages and the drawbacks are presented. The discussion pretends to give some criterions for selecting one of them in a specific application. This study was encouraged for importance motion information in recognition and imitation of gestures to be used in programming robots.

Keywords: motion detection, optical flow, correspondence, frequency techniques, artificial vision

1. INTRODUCTION

Motion estimation in artificial vision is one of the most studied areas. It is because motion can be used in much and diverse applications. Specifically, motion information has been used in tasks like: (a) video compression and codification; (b) satellite image processing; (c) civil and military applications of object tracking and autonomous navigation; (d) obstacle evasion in mobile robotic; (e) identification of anomalies using image processing in biological and medical images; (g) surveillance and supervision of places; (f) virtual reality and interfaces; (h) tridimensional structure recovery; (i) sport training, comparing behaviors respect a mathematical model; (j) fails localization and problems identification; and (k) speech recognition.

This work presents the main techniques in artificial vision for motion estimation. Techniques are classified in three groups: (1) differential techniques for optical flow estimation; (2) correspondence methods; and (3) methods in frequency domain. Each of them will be described in section two.

Discussion in section three provides criterions to choose the appropriate technique in a specific application. Finally, in section four conclusions and observations is done. Review of techniques in artificial vision for motion estimation was encouraged for importance motion information in gesture recognition. Recognition is the fist stage in robot programming using learning by demonstration for complex tasks.

2. TECHNIQUES FOR MOTION ESTIMATION

For motion estimation is necessary to have a couple of consecutive images at least. However, is possible to obtain a better estimation using more consecutive images, because it brings more information (Horn and Schunck, 1981). A better estimation means that motion estimation is closer to real motion (motion field).

There are techniques for motion estimation in 3D and 2D. This work is focused in second ones, which can be classify in:

a. Differential techniques to compute optical flow.
b. Correspondence methods.
c. Methods in frequency domain.

2.1. Differential Techniques to Compute Optical Flow

Optical flow is defined like a velocity field in the plane of the image that describes the motion of the pixels, and was defined rigorously by (Nakayama and Loomis, 1974).

Differential methods receive this name because they use first and second order spatiotemporal derivatives to calculate velocity vectors. Often, derivatives are computed using masks over brightness intensity of images.

Different techniques for optical flow estimation have been reported, and they use different assumptions. Nevertheless, there is one used for most of the techniques of optical flow and for all the differential ones, the constant of brightness hypothesis. This hypothesis was formulated by (Horn and Schunck, 1981), and establishes than in a short time interval, a same point keeps its brightness intensity from an image to the consecutive one, mathematically:

$$I_x v_x + I_y v_y + I_t = 0 \quad (1)$$

Where ***Ix***, ***Iy*** and ***It*** are the image derivatives in space (***x*** and ***y***) and time (***t***) respectively, and, $\boldsymbol{v_x}$ and $\boldsymbol{v_y}$ are the velocity vectors in ***x*** and ***y*** to be calculated. However, there is just one equation to get two unknowns, so, it is no possible to obtain a unique solution (aperture problem). It is just possible to determine the optical flow component in the gradient direction. This component is named normal optical flow and is defined by the equations (2).

$$v_{\perp x} = -\frac{I_t I_x}{I_x^2 + I_y^2}, \; v_{\perp y} = -\frac{I_t I_y}{I_x^2 + I_y^2} \qquad (2)$$

Another way computationally efficient to obtain normal optical flow was proposed by (Bradski and Davis, 2002). The idea is calculate a Motion History Image (MHI) using object segmentation (see figure 1). vectors pointing in the motion arm direction (orthogonally to the neighbor of the object can be obtained computing the gradient over MHI.).

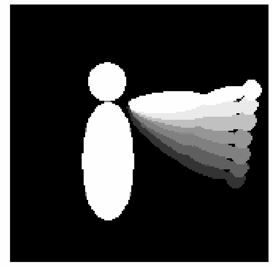

Fig. 1: Example of MHI for an arm motion sequence. From (Bradski and Davis, 2002)

Although normal optical flow is easy to calculate and it can be enough in some applications. Most of the time is necessary to obtain whole optical flow. To achieve this propose, new hypothesis or restrictions have to be included. For example, (Horn and Schunck, 1981) used equation (1) and a global term of smoothness, so, neighbors have similar motion. (Lucas and Kanade, 1981) uses the neighbors' information to fit the local condition defined by equation (1). Fitting is obtained using weighted least square minimization. (Nagel, 1983) was one of the firsts using second order derivatives. He also includes the smoothness condition used by Horn and Schuck,

Second order techniques are difficult to calculate in accurate way and often, they are affected by noise. For this reason, there are few second order methods and their estimations are lest accurate than the obtained for the first order ones (Horn and Schunck, 1981; Hildreth, 1984).

The main advantage of differential methods is that a dense vectorial field is obtained in a relatively simple manner. The main drawbacks are: (a) they are strongly affected by to noise; (b) they need a slow and smooth motion; (c) they need a continuous variation of motion in the image; (d) they are affected by occlusions and transparency; and (e) they can suffer of aperture problem.

2.2. Correspondence Methods

This method is the most easy to understand intuitively. It consists in localize different characteristics in successive images and determine their movement. Those characteristics can be edges, corners, lines, and so on.

The correspondence between characteristics in consecutive images will be optimal just using exhaustive search. However, this kind of search has high computational cost, for that reason, searching is usually limited inside a region. The size of the searching region will determine the behavior of the algorithm. Small search windows are good for slow motions and the number of positions to evaluate the correspondence is minor. Big search window will let greater motions, but it will be require high compute cost to evaluate more positions.

There are different correspondence motions and can be classify in:

Monoresolution: At the same time, this class can be sub-classify according to the characteristic used:
- Points association.
- Lines association.
- Block association.
- Group or Region association.
- Association of the position of neurons which follow shapes.

Multiresolution: Using hierarchical approximations.
- Hierarchical approximations for optical flow estimation.
- Multiresolution block correlation.

In the following paragraphs, each one of them will be described in detail.

Monoresolution:

- Points association: This technique can be defined like: Given ***n*** frames in different time instants, and ***m*** points in each frame. To associate each point in a frame to one point in the next frame. Two different points can not be associated to the same point in the next image. In this group stand out techniques like pixel matching, modify pixel matching and corner correspondence.

  Corners are especially important because they represent stable patterns along the time in the scene. They have a lot of information and they can be associated to geometric characteristics or to pixels' brightness intensity. There are many corners detectors. (Schmid and Mohr, 2000) made a comparison of corner detection algorithms and they found better results using the Harris' corner detector (Harris and Stephens, 1999).

- Lines association: These techniques try to extract the contours of objects, modeling them like line segments (see figure 2). Motion is determined by the displacement of the lines between consecutive images.

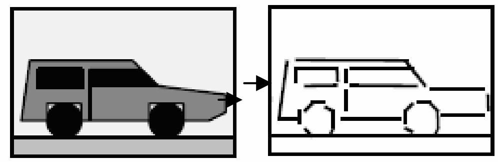

Fig. 2: Example of lines extraction. From (Zuloaga et al, 1997)

An interesting modification was used by (Fujiyoshi and Lipton, 1998) to determine relative motion between leg and torso of different people.

Walking or running people were recorded in a video sequence. The body of a person is approximated by a set of five lines forming a "start". Lines are obtained using a technique called skeletonization. Using those lines is possible to extract motion characteristics, for example, the relative angle between a vertical line passing by the centroid and the most rightness leg.

- Block association: It can be divided by:

<u>Block matching or Correlation of block with fixed size:</u> This method is one of the most popular for motion estimation because its low computational complexity. It uses blocks with fix size $N_1 x N_2$ and centered in the (n_1, n_2) position. The blocks are displaced inside a window in the following image looking for the best match position (n_1', n_2'). The best position is determined maximizing or minimizing some similarity criterion between blocks. Motion correspond to the subtraction $(\boldsymbol{n_1'}\text{-}\boldsymbol{n_1}, \boldsymbol{n_2'}\text{-}\boldsymbol{n_2})$. An important characteristic of this method is that all the pixels inside the block have assigned the same motion vector.

The most used criteria to find the best correspondence are: the minimum mean square error -MSE (Murat, 1995), the minimum mean absolute difference -MAD (Murat, 1995; Braskaran and Konstantinides, 1995), and the maximum matching pel count – MPL (Murat, 1995). However, you can use different criteria for the application.

<u>Block segmentation algorithms or Correlation of block with variable size</u>: The algorithms of block segmentation, in general, divide the

images in "seed" blocks, and then blocks can be divided again in smaller ones. The new division is required if some conditions or restrictions are not reached. Conditions can be, for example, the gray level dispersion (or color) inside the blocks (see figure 3). Then, the correspondence between blocks in two consecutive images is searched. Finally, the motion is determined by the displacement of the region centers between the matched regions.

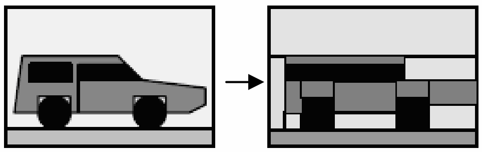

Fig. 3: Example of block segmentation. From (Zuloaga et al, 1997)

- Group or Region association: Like first step, searching process of regions require to group zones belonging to the same class. Regions are not limited to be blocks. When regions are matched using some searching algorithm, motion is determined by the centroid displacement.

Algorithms of group segmentation (clusters): These algorithms model the objects in the images like amorphous areas. Those areas are defined by characteristics in the image more or less uniform. Characteristics can be color, gray intensity, texture, etcetera (See figure 4). Motion is determined by the mass region displacement in successive images.

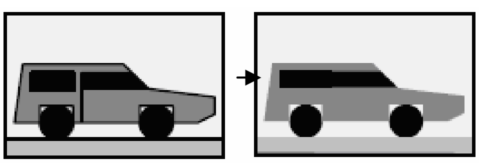

Fig. 4. Example of group segmentation. From (Zuloaga et al, 1997)

Centroid Tracking: In this method, background has to be captured before motion starting (Figure 5 left). More than one background images can be used to make this method more robust to light variations. When the object to be tracked comes into the environment (Figure 5 center), images are subtracted. Subtracted image is passed by a threshold and filter to obtain a binary image (Figure 5 right). Then, over the binary image, centroid and the maximum distance between this point and all the points in a circumference are computed. Circumference should include the entire object (See figure 6).

Fig. 5: Object Identification.

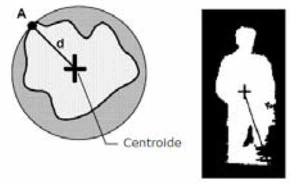

Fig. 6: Characteristics used for centroid tracking.

Temporal Image Averaging: This method is particularly useful when the background is unknown or changes during the sequence. It is a fast way to get motion information. Frames intensity is averaged according with equation (3).

$$g(t) = k * g(t-1) + (1-k) * f(t) \qquad (3)$$

Where ***g(t)*** is the current average, ***g(t-1)*** is the last average, and ***f(t)*** is the correspondence at current frame. A quiet object for a while has a black representation.

- Association of the position of neurons which follow shapes: PhD thesis work developed by (Flóres, 2001) is very interesting; it is because he uses an artificial intelligence technique like neural network to solve the motion detection problem. In addition, it lets to model no rigid objects.

Flóres used a kind of networks named Growing Neural Gas –GNG (Fritzke, 1995). GNG networks preserve topology of an input space (object). Moreover, GAG networks let dynamic object modeling in time (morphologic variations). In this way, motion of the object is represented by the trajectory of the neurons' position along the map. Since neurons' position is already known, correspondence problem is avoided. Correspondence process is computationally demand in most of the object tracking techniques. However, shape training required for this method is not too fast, especially for first frame in the sequence.

Multiresolution or hierarchical approximations: The idea of these kinds of algorithms is to use several resolutions. First, low resolution estimation is performed, and next, the estimation is refined using a higher resolution.

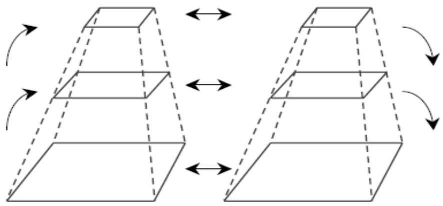

Fig. 7: Multiresolution algorithms

- Hierarchical approximations for optical flow estimation: (Burt et al., 1989; Bergen et al., 1990) used correlation in multiple scales to calculate optical flow for several simultaneous motions. However, it can be problematic to share information between different scales; especially when the aperture problem is present. Figure 8 is an example where the big square is moving upwards and the scale used (Small Square) does not let determine the motion. Even using highest scales can not be possible to determine it.

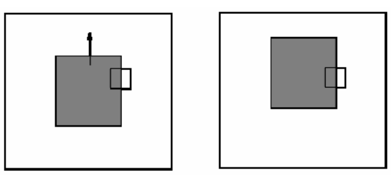

Fig. 8: Aperture problem example

The main problems of this method is computational cost and to determine the size window to perform the correlation process.

- Multiresolution Block correlation: This method can be classified in two big groups: The first one uses block correlation in a pyramid Gaussian (Burt and Adelson, 1983). Main idea is that fast motions can be estimated using smoothed and subsampled versions of the images (LeQuang and Zaccarin, 1995). Second group corresponds to multigrid that uses block correlation in quadtree structure. A mosaic of spatial blocks with different sizes is created. Different sizes permit that the analyzed region can be adapted to the underlying data (Dufaux and Moscheni, 1995).

In general, correspondence methods are less sensible to noise. It is because they use more data to estimate motion than other ones do not. Vectorial field using this technique is not much dense, although it can be advantageous in some applications.

2.3. Methods in Frequency Domain.

Methods in frequency domain can be divided in two classes: phase based methods and energy based methods.

Phase based methods: These methods use the fact than movement in spatial domain produce a phase change in frequency domain (Papoulis, 1984); mathematically it is described by equation (4). One example of this kind of technique was proposed by (Fleet and Jepson, 1990). Fleet and Jepson applied a gradient based technique to the phase component. Phase component is the output from a set of directional pass-band filters tuned to different velocities.

$$f(x-\Delta) \rightarrow F(w)e^{-iw\Delta} \qquad (4)$$

Where $F(w)$ is the Fourier transform of $f(x)$.

Energy based methods: These methods work with the energy distribution in the frequency domain and they consider the motion in their spatio-temporal orientations. Energy based methods are broadly accepted by the neuroscience community like models of some areas in the path motion. Particularly, tuned Gabor filters' outputs correspond to neurons' responses in the primary visual area –V1 (Adelson and Bergen, 1985). Combination of Gabor filters' responses tuned to different spatio-temporal orientation can be used in speed and direction motion estimation (Adelson and Bergen, 1985; Heeger, 1987). This process is performed in the middle temporal area – MT.

Methods in frequency domain present some remarkable advantages respect many methods in spatial domain. In first place, methods in frequency domain are lest sensible to global illumination changes and they are quite robust to noise. By the other hand, they can detect motion of points with random shapes that can not be possible detect using correspondence or differential techniques.

3. DISCUSSION

Techniques used to recover motion of objects from a video sequence in artificial vision can be divided in three big classes: differential techniques for optical flow estimation, correspondence methods or displacement based methods, and techniques in frequency domain.

Applications that require dense vectorial field should be consider using differential techniques to optical flow estimation. Do not lose of mind the computation time when you chose one of them. Moreover, it is important to verify that restrictions or assumptions used for the elected optical flow technique to recover motion information are fulfilled (approximately at less).

When robustness to noise is required, correspondence methods or methods in frequency domain must be considered. However, the formers are more robust than the phase based methods (Mandushi and Mian, 1993). Nevertheless, the correspondence methods' results can be limited for the elected window size. As differential techniques for optical flow estimation, correspondence methods are sensible to illumination changes. By the other hand, if the application requires a dense vectorial field, correspondence methods is not a good option.

If the application requires both: global illumination changes and noise robustness, it is recommended to use techniques in frequency domain. Specifically, methods in frequency domain perform a normalization operation named whitening. This operation brings robustness to global illumination changes.

In artificial vision is already looking for techniques for motion estimation with robustness to: noise in sensors' reading, illumination changes, noise in the recorded images, occlusions, transparencies, and so on. At the same time, it is desirable that those techniques have enough low computational cost and time to be used in real time applications.

4. CONCLUSIONS

A review of motion estimation techniques and their classification was presented. Those techniques perform the estimation from sequence of images.

It is clear that the election of one of the techniques for motion estimation depends of the application where it will be applied. The environment conditions where the task will be performed must be considered as well.

By the other hand, in spite of a large and diverse amount of artificial vision techniques, there is not a predominant one to solve the motion estimation problem. Nevertheless, optical flow looks like the most studied ones; and they are used in many artificial vision applications. In fact, optical flow techniques are studied jet to achieve a better motion field estimation, to reduce computation time. Those characteristics will allow that optical flow techniques can be used in more real time application.

In the bibliographic review is remarkable that was found few work using artificial intelligence methods for motion estimation. Nowadays, artificial intelligence techniques have be used in many and diverse fields; technology advances encourage their application. Therefore, it is an interesting research area in motion estimation. For this purpose can be used any: neural networks, fuzzy and ants systems, among others.

Review of artificial vision techniques for motion estimation was motivated for the interest of the research group of Perception and Intelligent Systems in the Universidad del Valle in programming robots by demonstration. Motion is considered a fundamental characteristic to achieve recognition and subsequent imitation. However, the group is especially interesting in bioinspired techniques for motion estimation. It is because biological systems have simple and successful strategies to solve different problems in natural environments.

ACKNOWLEDGMENTS

We thanks to '*Colciencias*' program "Support to PhD Studies", to the Universidad del Valle and the professor José Santos-Victor from the Institute of Systems and Robotic in Lisbon-Portugal, for the support to this project.

REFERENCIAS

Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, **2(2)**, 284-299.

Bergen, J. R. , Burt, P. J., Hingorani, R. and Peleg, S. (1990). Computing two motions from three frames. *Proc. 3rd Int. Conf. on Computer Vision*, 27-32.

Bhaskaran, V. and Konstantinides, K. (1995). *Image and Video Compression Standards.* Kluwer Academic Publishers.

Bradski, G. R. and Davis, J. W. (2002). Motion segmentation and pose recognition with motion history gradients. *Matching Vision* and Applications, 13, pp. 174-184, 2002.

Burt, P. J. and Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, **31(4)**, 532-540.

Burt, P. J., Bergen, J. R., Hingorani, R., Kolcznski, R. J., Lee, W. A., Leung, A., Lubin, J. and Shvaytser, J. (1989). Object tracking with a moving camera, an application of dynamic motion analysis. In *IEEE Workshop on Visual Motion,* 2-12, Irvine, CA.

Dufaux, F. and Moscheni, F. (1995). Motion estimation techniques for digital TV: A review and a new contribution. *Proc. of the IEEE*, **83**, 858-876.

Fleet, D. J. and Jepson, A. D. (1990). Computation of component image velocity from local phase information. *International Journal of Computer Vision*, **5**, 77-104.

Flórez, F. (2001). Modelo de representación y procesamiento de movimiento para diseño de arquitecturas de tiempo real especializadas. In: *Departamento de Tecnología Informática y Computacional*. Alicante: Universidad de Alicante, 169.

Fritzke, B. (1995). A Growing Neural Gas Network Learns Topologies. In: *Advantages in Neural Information Processing Systems 7*, G. Tesauro, D. S. Touretzky, and T. K. Leen, Eds. Cambridge, Mass.: MIT Press.

Fujiyoshi, H. and Lipton, A. (1998). Real-time human motion analysis by image skeletonization. *IEEE Workshop on Applications of Computer Vision*

Harris, C. and Stephens, M. (1999). A combined corner and edge detector. *Alvey Vis Conf*, 147–151.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America*, **4(8)**, 1455-1471.

Hildreth, E. (1984). Computations Underlying the Measurement of Visual Motion. *Artifitial Intelligence*, **23**, 309-354.

Horn, B. and Schunck, B. (1981). Determining Optical Flow. *Artifitial Intelligence*, **17**, 185-203.

LeQuang, D. and Zaccarin, A. (1995). Object-Oriented Coding Using Successive Motion Field Segmentation and Estimation. *Proc. of ICIP*, **1**, 207-210.

Lucas, B. and B. Kanade, B. (1981). An iterative image registration technique with an application to stereo vision. *Proc. DARPA Image Understanding Workshop*, 121-130.

Manduchi, R., and Mian, G.A. (1993). Accuracy analysis for correlation-based image registration algorithms. *Proc. IEEE-ISCAS*, Chicago, IL, USA, 834–837

Murat, A. (1995). *Digital Video Processing*, Prentice Hall.

Nagel, H. H. (1983). Displacement vectors derived from second order intensity variations in image sequences. *Comput. Graph. Image Process*, **21**, 85-117.

Nakayama, K and Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception*, **3**, 63-80.

Papuolis, A. (1984). *Signal Analysis*, McGraw-Hill.

Schmid, C., Mohr, R. and Bauckhage, C. (2000). Evaluation of interest point detectors. *I J Comp Vis*, 37(2), 151–172.

Zuloaga, A., Martin, J.L., Nozal, L.L (1997). Determinación del movimiento a partir de secuencias de imágenes. Disponible en: http://www.geocities.com/aitzol.geo/docu005.htm