Fuzzy Approach of Motion Estimation

Frédéric Comby, Olivier Strauss, Marie-Jose Aldon
LIRMM, UMR CNRS/Universite Montpellier II,
161 rue Ada, 34392 Montpellier cedex 5, France

Abstract— **This paper deals with main apparent motion estimation using a fuzzy representation of pixel's gray level. Rough histograms and possibility theory are used to provide an accurate estimation of the motion. The use of these theories allows us to partly cope with classical motion estimation method limitations. This article explains in details the estimation process and also presents some illustratives examples.**

I. Introduction

Signal processing and computer vision have been in full expansion for the past ten years. Motion estimation from a video sequence is a key issue in image processing and computer vision. It has many applications in object tracking, egomotion estimation of a mobile robot, or mosaicing. This article aims at finding the main apparent motion in a video sequence. This field has already been thoroughly studied and many methods have been tested. These methods gather together correlation, optical flow and feature-based methods. However, these approaches rely on strong hypothesis that have to be frequently transgressed - thus limiting their reliability.

A state of art proves that few works focus on the improvement of fuzzy and possibility theories with regard to motion estimation. This paper offers to use a fuzzy model of the image to represent data. Moreover, possibility and rough histograms theories allow to partly cope with inherent problems of classical motion estimation methods.

After this first introductory chapter, the principles of apparent motion are defined in chapter two. Classical methods and their drawbacks are then briefly presented. The third chapter introduces a new approach based on a fuzzy modeling of pixels. The motion estimation method is also set out. Chapter four unfolds some results. Eventually, the last chapter concludes on the contribution brought by the method and raises new perspectives

II. Apparent motion

A. Definitions

In a static scene, the movement of a camera entails an apparent motion on the sequence it acquires. This movement results in a variation of the environment point's projection on the camera's focal plane. However, phenomena such as occlusions, moving objects in the static scene, or global illumination variation can be interpreted as apparent motion. Most classical methods aim at finding the main apparent motion linked to the camera's displacement.

In a 3-D environment, the main apparent motion on image space due to the motion of the camera can generally be modelled according to six parameters (three translations and three rotations). However, Bouthemy in [1] has shown that under small displacement assumption, a four parameter model is more robust. The pan and tilt rotations of the camera can be related to translations in the focal plane. The motion is then characterized by two translations on each axes of the image, one rotation around the normal to the focal plane and a scaling factor.

Classical methods for motion estimation can be divided in three categories: feature, correlation and optical flow-based methods.

B. Feature-based methods

It consists in finding specific points (corners, image edges) with transformation-invariant properties (such as translations, rotations, change of global brightness, etc.). These points are called features. Once these features have been extracted, a matching process finds the shift between the same feature in two successive images. One of the drawback is the arbitrary choice of feature sent to the matching process. If - due to an occlusion - an interesting point vanishes in the second image, the motion estimation may fail or its robustness decrease.

C. Optical flow-based methods

These methods are among the most studied [2], [3], [4]. Let $E(x,y,t)$ be the irradiance image. Motion estimation methods based on optical flow computation link the spatio-temporal gradient of $E(x,y,t)$, using the constraint equation:

$$\nabla E_{(x,y)}[\begin{smallmatrix} u \\ v \end{smallmatrix}] + E_t = \xi, \tag{1}$$

where $\nabla E_{(x,y)}$ is the spatial gradient, and E_t the temporal gradient of the irradiance image; $(u,v) = (\frac{dx}{dt}, \frac{dy}{dt})$ is the projection of the 3-D velocity field in the focal plane, and ξ is the global brightness variation.

As this computation is based on brightness continuity, only small displacements can be estimated. Moreover, this approach is based on two antagonist assumptions. On the one hand, the image must be sufficiently textured for the motion to be visible. (If $\nabla E_{(x,y)}$ and E_t are zero, no motion can be estimated.) On the other hand, the computation of local gradients $\nabla E_{(x,y)}$ and E_t is made through a low-pass filter, which requires a low texture of the image.

Reducing the influence of these constraints is possible. For example, Bouthemy in [1] offers to use robust statistics in order to cope with pixels not fulfilling (1). He also suggests to use a multi-scale process so as to allow large displacements estimation.

D. Correlation-based methods

These methods split up in two approaches: parametric and non-parametric statistical-based methods [5]. In both cases, a pixel to be matched becomes the center of a small window of pixels in the first image. This window is compared with similarly sized regions in the second image. Then parametric or non parametric statistical-based matching metrics are used to provide a numerical measure of similarity between these two windows.

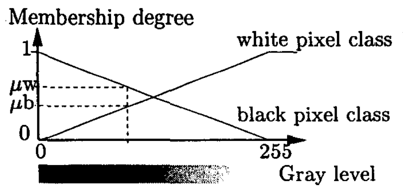


Fig. 1. Fuzzy membership function of pixels' class.

These methods require a motion model. The parametric space of the motion is discretized, and the couple of parameters obtaining the best similarity measure is said to square with the main motion. However, the precision of the estimation is limited by the sampling interval of the motion parameter's space. Moreover, correlation-based methods do not correctly handle motion model containing rotations because it involves non-integer displacements. Finally, it is necessary to set an arbitrary threshold to discard meaningless estimation. The arbitrary nature of this thresholding lowers the estimation's robustness.

III. New approach

The approach presented here has certain similarities to the methods described above. as with correlation-based methods, a discretization of the parametric space is required. Moreover, as optical flow-based methods, it links spatio-temporal variation of the irradiance image.

A. *basic concept*

The camera displacement entails a movement of the environment points projections. Feature based-methods assume that the peculiarity of local properties is conserved. Those based on correlation admit that local patterns are retained through the image sequence. As for optical flow-based methods, the assumption is about the local brightness distribution retaining. The approach presented here is less restricting because based on a very simple hypothesis. Instead of comparing the pixel's intensities, the process rather tests if a pixel (i, j) of picture one and a pixel (i', j') of picture two belong to the same class.

The hypothesis can then be written as: a white pixel is unlikely to become a black pixel and vice-versa; but a gray pixel can become a black or white pixel.

In order to simplify this reasoning scheme, the gray level-space is split in two dual classes: black and white pixels. Classes are fuzzified (Fig. 1) to avoid instability around the classes' boundaries. These two fuzzy dualistic classes model minimizes the pixel representation issue apriority. Moreover, the use of classes is known to enhance the robustness of a process.

Except for the camera's motion, other causes may induce a pixel's intensity change: a global brightness variation, a moving object in front of the camera, or a change of a surface's reflective properties due to a variation of orientation.

The issue of global brightness variation is partly settled with the fuzzy dual model of pixel's gray level. Indeed, a small variation of a pixel's intensity will only slightly change its membership degrees to black and white class. The two other problems will be treated later on in this paper.

B. *Motion Characterization using the pixel's fuzzy representation*

Let's assume that the model of the motion to be estimated is a three parameters' model: two translations (T_x, T_y) and a rotation (θ) (the zoom factor issue hasn't been tackled with yet). In order to compute the motion estimation, the parametric space is discretized. The estimation process amounts to finding the most likely displacement, in the parametric space defined above.

If a pixel P_1 located in (i, j) on the first image is transformed with a (T_x, T_y, θ) motion, it will be located in (i', j') on the second image and referred to as P_1' Fig 2.

The likelihood of a pixel's displacement is defined as:

Axiom 1: The (T_x, T_y, θ) displacement that transforms P_1 into P_1' is likely if both P_1 and P_1' belong to the same class. That is P_1 **and** P_1' are white pixels, or P_1 **and** P_1' are black pixels.

Because of the fuzzy modeling of pixels' classes, a pixel is characterized by two membership values. The likelihood of a displacement is then dually evaluated. The eventuality for a pixel P_1 of the first image to be transformed in P_1' in the second image is represented by:

$$\begin{aligned} & \mathrm{Vote}_{pro}(P_1'; P_1) \\ = \; & \max(\min(\mu_w(P_1); \mu_w(P_1')); \min(\mu_b(P_1); \mu_b(P_1'))), \end{aligned} \tag{2}$$

where $\mu_w(P)$ (resp. $\mu_b(P)$) is the pixel P membership degree to the white pixel's class (resp. black pixel's class). The dual quantity, representing the impossibility for a pixel to move from P_1 to P_1', is evaluated by:

$$\begin{aligned} & \mathrm{Vote}_{con}(P_1'; P_1) \\ = \; & \max(\min(\mu_w(P_1); \mu_b(P_1')); \min(\mu_b(P_1); \mu_w(P_1'))). \end{aligned} \tag{3}$$

If the gray level of pixels P_1 and P_1' are known precisely, then:

$$\mathrm{Vote}_{con} = 1 - \mathrm{Vote}_{pro}. \tag{4}$$

This method allows to partly cope with the problems linked to the image's border. Indeed, if a displacement (T_x, T_y, θ) brings the pixel P_1' out of the second image, then the gray level of this pixel is unknown. The lack of information about P_1' gray level is taken into account by setting $\mu_b(P_1')$ and $\mu_w(P_1')$ to 1.

C. *Link with the motion model*

The motion estimation process has to take into account the imprecision of all the computation process.

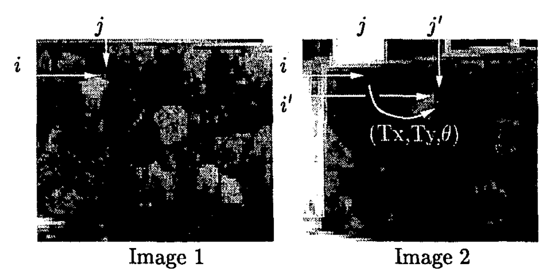


Fig. 2. Pixel's displacement under (T_x, T_y, θ) movement.

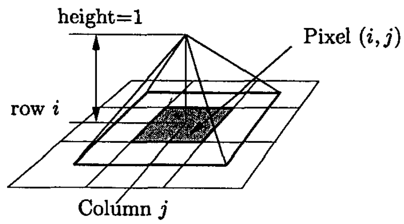


Fig. 3. Fuzzy representation of a pixel.

A pixel has to be considered as an imprecise quantity because of the picture digitalization. Its representation is then at least a crisp box or a fuzzy box, resulting from the cartesian product of two crisps or fuzzy intervals in 2-D. Hereinafter, the fuzzy representation of pixels will be used. The pixel's membership function's variation will be linear, so as to remain as neutral as possible - no assumption on data distribution in the interval (Fig. 3).

The parametric space discretization also brings imprecision on the parameters of the motion model. This imprecision is modelized by a fuzzy representation of the parametric space partition. A cell of this partition is then characterized by its kernel - value used in classical partition - and its support - linked to the sampling interval. The transformation that brings P_1 from the first image to P_1' from the second one can be re-written as:

$$P_1' = \mathcal{R} \times P_1 + \mathcal{T}, \tag{5}$$

where $\mathcal{R}$ is the ensemblist transformation corresponding to an imprecise rotation of angle θ_0, with a support defined by $[\theta_0 - \Delta\theta, \theta_0 + \Delta\theta]$, and $\mathcal{T}$ is the ensemblist transformation corresponding to a $[T_{x0}, T_{y0}]$ translation, with supports defined by $[T_{x0} - \Delta T_x, T_{x0} + \Delta T_x]$ and $[T_{y0} - \Delta T_y, T_{y0} + \Delta T_y]$. $\Delta\theta$, ΔT_x and ΔT_y are linked to the parametric space discretization.

As P_1' results from the ensemblist transformation of P_1, it is also an imprecise quantity. Its representation covers many pixels P_2 of the second image Fig. 4.

The focus is now put on how a pixel P_1' will vote for a displacement (T_x, T_y, θ). With a possibilistic reasoning, the compatibility between the two imprecise quantities involves an average over all pixels covered by P_1'. The fuzzy approach involves two computations: the worst case, where all pixels partially or totally covered by P_1' are taken into account; and the best case, where only totally covered pixels are taken into account.

If the possibility theory's terminology is used, then the following claims can be written:

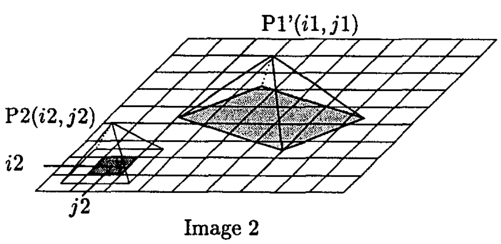


Fig. 4. P_1' covering many pixels P_2 from picture 2

Axiom 2: Pixel P_1 will vote in a possible way for (T_x, T_y, θ) if, $\exists P_2$ satisfying $P_2 \cap P_1' \neq \emptyset$ and gray level of P_2 and P_1 belong to the same class.

Axiom 3: Pixel P_1 will necessary vote for (T_x, T_y, θ) if, $\forall P_2 \subset P_1'$, gray level of P_2 and P_1 do not belong to different classes.

To estimate the compatibility between quantities P_1' and P_2 in terms of their degree of inclusion and intersection, the possibility and necessity measures defined in [6] are used:

$$\begin{aligned} \text{Possible Vote} &= \Pi(P_1'; P_2), \\ \text{Necessary Vote} &= N(P_1'; P_2), \end{aligned} \tag{6}$$

where $\Pi(P_1; P_2)$ (resp.$N(P_1; P_2)$) is the conditional possibility measure (resp. necessity measure) of P_1' given P_2. The sup − min composition rule [7] is then used to merge (2), (3) and (6), which gives:

$$\begin{aligned} & Vote_{Sup}(T_x, T_y, \theta)_{P_1} \\ = \;& \underset{P_2 \in \text{Image 2}}{Sup} \min(\Pi(P_1'; P_2); \text{Vote}_{pro}(P_1; P_2)), \end{aligned} \tag{7}$$

$$\begin{aligned} & Vote_{Inf}(T_x, T_y, \theta)_{P_1} \\ = \;& 1 - \underset{P_2 \in \text{Image 2}}{Sup} \min(N(P_1'; P_2); \text{Vote}_{con}(P_1; P_2)). \end{aligned} \tag{8}$$

These votes are performed for each pixel and each combination of the parametric space. Then, they are collected in two accumulators associated to the fuzzy partition of the parametric space. These accumulators are respectively called lower accumulator (polling the $Vote_{Inf}$) and upper accumulator (polling the $Vote_{Sup}$). These accumulators, associated with the fuzzy partition, define a rough histogram. This concept is fully presented in [8].

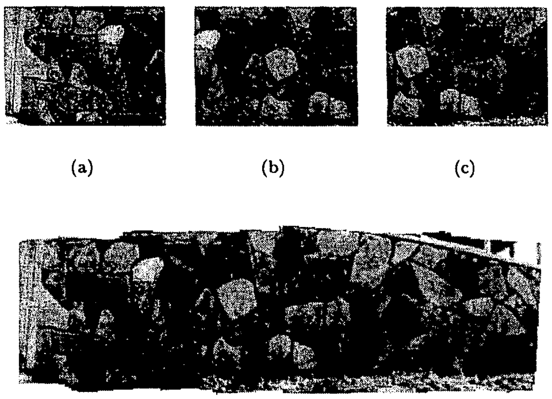
Once the rough histogram has been computed, it defines a framing of the apparent motion's density of probability. The next step aims at finding the main mode of this rough histogram, which is associated to the main apparent motion.

D. Motion estimation

Finding the main mode in a classical histogram consists in finding the cell whose accumulator is maximal. However, the localization's precision of this mode depends on the width of the histogram's bins. Using rough histograms entails a kind of natural interpolation between the discrete values of the partition [9], thus improving the localization's precision.

The full main mode estimation method is presented in [8], and only the basis are explained here. Searching the main mode in a rough histogram amounts to searching the position of a crisp or fuzzy interval ϕ of precision Γ, polling a locally or globally maximum of votes. The number of votes purporting to this interval ϕ has then to be estimated.

This estimation is achieved by transposing the imprecise number of votes towards the interval ϕ. This belief transfer uses the pignistic probability defined in [10] and [11]. Then, an upper bound (resp. lower bound) of the number of votes embodied in ϕ called $\overline{Nb(\Phi)}$ (resp. $\underline{Nb(\Phi)}$) is computed with the upper accumulator (resp. lower accumulator) of the rough histogram.



(a) (b) (c)

(d)

Fig. 5. Images number 1 (a), 35 (b) and 54 (c) from the 120-images sequence of the wall. The image (d) represents the resulting mosaic.

Up to now, the main mode is searched as the position of ϕ maximizing $\overline{Nb(\Phi)}+\underline{Nb(\Phi)}$ (kind of average probability). Other formulas, derived from statistical reasonning with imprecise probabilities defined in [12], have been successfully tested. This part is still being improved today.

As for the problems linked to an occlusion or a change of a surface's reflective properties due to a variation of orientation, if the pixels altered by their effects are in the minority, the main mode estimation will not be affected. Thus, the estimation of the main apparent motion partly copes with such problems.

IV. Results

The mosaic presented in Fig.5 has been run out on a 120-images sequence of a stone wall. The motion of the camera was mainly composed of a horizontal translation. Meanwhile, the motion also includes small rotations around the normal to the focal plane and vertical translations. The mosaic is created by estimating the motion (T_x, T_y, θ) between two consecutive images and by superposing them on the resulting picture.

The mosaic image emphasizes the estimation accuracy since no visible disruptions occur on the surface wall.

The three parameters model used for the estimation process (T_x, T_y, θ) is an approximation of the real motion. Actually, it also includes zoom and other rotations - considered as parasite motions. The motion estimation process does not appear to be disturbed by this approximation, thus, as far as the model is concerned, the motion estimation is robust.

However, the fact that the camera was not moving in a plane parallel to the wall brings small distortions on the resulting mosaic.

During the sequence acquisition, the operator kept varying the lens aperture so as to produce variations in the global brightness. As such, the gray level of the projection of a 3-D environment specific point can be different between two consecutive images. This does not interfere with the motion estimation process. Thus, as far as data contamination is concerned, the estimation is also shown to be robust.

V. Conclusion

A new method of motion estimation has been presented in this paper. This estimation is based on a fuzzy modelling of pixel's gray level and uses both rough histograms and possibility theories. This method allows to partially cope with classical issues in motion estimation process, such as small displacement assumption and texture-linked constraints. A motion model and a discretization of the motion parameters space is needed. However, using rough histograms involves a somewhat natural interpolation, and leads to an estimation less sensitive to the parameters space discretization.

Moreover, as far as data contamination is concerned, using a fuzzy model of pixel's gray level appears to be robust. Actually, a variation of global illumination does not cause the estimation process failure.

Another positive point is that the method gives an estimation of the motion as well as a confidence measure of this estimation. Indeed, the gap between the lower and upper accumulators is linked to this confidence measure. The more the accumulators remain apart from each other, the less significant the estimation is. Unlike correlation-based methods, the thresholding for keeping the estimation relevant, is set by data.

However, this motion estimation process has drawbacks. First, the computing time is quite long since loops are intertwined in the algorithm, but improvements are in progress.

Then, the overlapping rate between two images has to be at least of 75 per cent, for fear of not guaranteeing the "main motion" estimation's success. Eventually, defining an arbitrary research area for the motion model parameters is necessary and entails a limitation of the estimated motion.

The next step of the motion estimation process will consist in improving the motion model with the incorporation of a scaling factor. The estimation process will also be tested on image stabilization issues.

References

[1] J. M. Odobez and P. Bouthemy. Robust multi-resolution estimation of parametric motion models applied to complex scenes. Pi 788, IRISA, 1994.

[2] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[3] Amar Mitiche and Patrick Bouthemy. Computation and analysis of image motion: A synopsis of current problems and methods. *International Journal of Computer Vision*, 19(1):29–55, 1996.

[4] D. J. Fleet J. L. Barron and S. S. Beauchemin. Systems and experiment: Performance of optical flow techniques. *International Journal of Computer Vision*, 12:1:43–77, 1994.

[5] M. Bennamoun J. Banks and P. Corke. Fast and robust stereo matching algorithms for mining automation. *JAF 97*, 1:139–149, November 1997.

[6] D. Dubois, H. Prade, and C. Testemale. Weighted fuzzy pattern matching. *Fuzzy sets and systems*, 28:313–331, May 1988.

[7] D. Dubois and H. Prade. *Possibility Theory An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.

[8] F. Comby, O. Strauss, and M.-J. Aldon. Possibility theory and rough histograms for motion estimation in a video sequence. *International Workshop on Visual Form IWVF4*, May 28-30 2001.

[9] O. Strauss, F. Comby, and M.-J. Aldon. Rough histograms for robust statistics. 2:688–691, September 2000.

[10] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–243, 1994.

[11] T. Denoeux. Reasoning with imprecise belief structures. Technical Report 97/44, Universite de Technologie de Compiegne, Heudiasyc Laboratory, 1997.

[12] P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, london, 1991.