Template Tracking using Generalized Modus Tollens

Olivier Strauss, Frédéric Comby

LIRMM, Université Montpellier II, 161, rue Ada,
34392 Montpellier Cedex 5, France
e-mail: Olivier.Strauss@lirmm.fr

***Abstract*: This article describes an original method to track 2D templates in a video sequence. This approach is based on robust modelling of the interaction between pattern movement and grey level variations in the image using fuzzy bimodal rules. Movement estimation is performed by inverting these rules and polling the results in a quasi-continuous histogram. Some experimental results are given to illustrate the performance of this algorithm.**

1. Introduction.

Template tracking has various applications within the scope of robotics (visual servoing), medical analysis, surveillance, human-computer interaction, video databases and 3D-reconstruction (stereo vision, structured light), to name but a few.

Template tracking generally consists of finding the position of the projection of an object in the image plane while this object is moving in front of the camera. The tracking problem is to find the *best* set of parameter values describing the motion of the target through a sequence.

When an object moves in front of a camera, the retina illumination distribution is modified in a coherent manner, called *apparent motion*, which is the signature of projected motion. Template tracking relies on an analysis of apparent motion to compute correspondence through a sequence of images. However, other phenomena such as partial or total occlusion, lighting variations, changes of viewpoint, reflectance, sampling and digitalization can modify the illumination in a quite coherent manner which is only slightly or not related to projected movement. Template tracking techniques can be divided in three main groups: feature-based, area-based and optical flow-based tracking.

Feature tracking consists of matching local primitives extracted from images such as segments of straight lines or curves, characteristic points, blocks or regions. Matching is based on minimizing statistical distances between attributes of the primitives like brightness, local contrast, shape, relative position, length, curvature, etc. One major advantage of these methods is their robustness relative to unexpected movement or variations of illumination such as global brightness, occlusion, pose variation etc. However, in natural images, strong visual features can sometimes seldom be found unless artificial beacons are placed in the environment. The fact that the template to be tracked requires a model is one of the major drawbacks of this kind of technique.

Area-based matching techniques use statistical distance measures directly on the grey-level. The tracked template is supposed to be known by its projection on a limited region of the image plane. It is generally assumed that the sought after displacement of the tracked template belongs to a known parametrized motion function. No explicit modeling of the template is needed. When using robust statistical distances, these methods appear to be very robust with respect to random global or local lighting modifications. However, due to their computational complexity, real-time application of area-based techniques is generally limited to translation, i.e. the higher the dimensional space the higher the computational cost.

More recently, a third efficient framework was proposed that uses an optical-flow like differential representation of the link between the target movement and grey level variations. In this framework, the illumination of the target template is supposed to be to be known in the first image. A learning process computes the relation between temporal illumination variations and parametric apparent movement. This learning process is able to directly estimate the projected movement by comparing grey level values of the target template with grey level values of a predicted region. The method we propose in this paper is in line with this kind of framework.

2. Differential tracking.

This part is a simplified version of the explanations given in [1] and [2]. The reader should refer to these publications for a more detailed explanation.

A. Principle.

A sequence of images can be viewed as a function $I(\mathbf{x},t)$ of the brightness at the pixel location $\mathbf{x}^t=(x,y)$ at time t. The target to be tracked is defined, on the first image, with a list of n pixels located at $\mathbf{x_i}^t=(x_i, y_i)_{i=1\ldots n}$. The illumination of this reference template is $I(\mathbf{x_i},0)$.

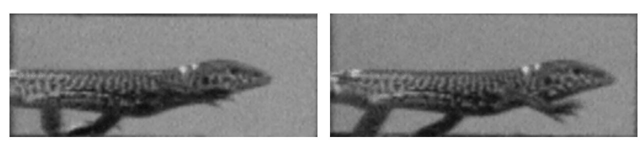


Figure 1: Two consecutive images of a running lizard

The motion of the target induces a variation in the grey level distribution of the template (Fig. 1). Let us assume that the motion of each pixel of the target region can be completely described by a function f and a parameter vector $\mathbf{ß}^t=(ß_1, \ldots, ß_m)$. Given two images at time 0 and time t, and assuming that there are no changes in the illumination of the target, the grey level distribution of the reference template follows:

$$I(f(\mathbf{x_i},\mathbf{ß(t)}),t) = I(\mathbf{x_i},0) + \varepsilon_i \qquad (1)$$

where ε_i represents unexpected variations in the grey level distribution due to changes in the target pose, deformation or partial occlusion. If the error ε_i has a known distribution, then the motion parameter $\mathbf{ß(t)}$ of the target can be estimated by a proper minimization of $(I(f(\mathbf{x_i},\mathbf{ß(t)}),t) - I(\mathbf{x_i},0))$.

B. From motion to grey level variation.

Let us now suppose that the projected motion is small enough (or the acquisition frequency is high enough) that frame to frame disparity is low. Then a first order differentiation of equation (1) allows us to compute grey level variations at location $\mathbf{x_i}$ when the parametric movement $\boldsymbol{\partial\beta}$ is known:

$$\partial I(\mathbf{x_i}) = M_i.\boldsymbol{\partial\beta} + O_i(\boldsymbol{\partial\beta}) \quad (2)$$

where M_i is the partial derivative of I with respect to $\boldsymbol{\beta}$ at the location $\mathbf{x_i}$. M_i can be easily computed from image differences and the parametrized function f. $O_i(\boldsymbol{\partial\beta})$ includes higher order terms and modeling errors. Equation 2 is called *optical flow constraint*. It is the basis of differential tracking methods.

C. From grey level variation to motion.

The template tracking principle proposed by Hager [1] consists of inverting equation (2) using the partial derivative of the image with respect to $\boldsymbol{\beta}$ and time parameter. Let ∂t be the time delay between the acquisition of the first and the second image. At time ∂t, at location $\mathbf{x_i}$ the illumination function is $I(\mathbf{x_i},\partial t)$. The illumination variation at location $\mathbf{x_i}$ can be linked to parameter motion $\boldsymbol{\partial\beta}$ using equation (2). Expression of equation (2) for each pixel of the reference template gives n equations with m unknowns that can be easily computed by a statistical method minimizing a proper distance between $\partial I(\mathbf{x_i})$ and $M_i.\boldsymbol{\partial\beta}$ (i.e. finding $\boldsymbol{\partial\beta}$ such that $I(f(\mathbf{x_i},\boldsymbol{\partial\beta}),\partial t) \approx I(\mathbf{x_i},0)$).

The advantage of this method (especially compared to area-based matching methods) is that no coverage of the parameter space is needed. It allows real time implementation because of the low-complexity of this algorithm. However the use of the standard least square solution to compute motion from grey level variations is not robust enough since this kind of expression of the problem with is often ill-conditioned. In [3], M. La Cascia et al. successfully combined regularization techniques with robust non-linear least square methods.

In a recent article [2], F. Jurie presents an efficient method to avoid this hazardous least-square inversion. Let us rewrite the m equations in a matrix format:

$$\boldsymbol{\partial I} = \mathbf{M}.\boldsymbol{\partial\beta} + \mathbf{O}(\boldsymbol{\partial\beta}) \quad (3)$$

with $\boldsymbol{\partial I}=[\partial I(\mathbf{x_1})\ldots\partial I(\mathbf{x_n})]^T$, $\mathbf{M}=[M_1\ldots M_n]^T$ and $\mathbf{O}(\boldsymbol{\partial\beta})=[O_1(\boldsymbol{\partial\beta})\ldots O_n(\boldsymbol{\partial\beta})]^T$. $\mathbf{M}$ is a (n,m) matrix. Inversion of (3) gives:

$$\boldsymbol{\partial\beta} = \mathbf{H}.\boldsymbol{\partial I} + \mathbf{O}(\boldsymbol{\partial I}) \quad (4)$$

The *hyperplane approximation* method involves directly learning matrix $\mathbf{H}$ from synthetic motion of the reference template in the initial image. When matrix $\mathbf{H}$ has been estimated, the motion can be directly computed from equation (4). Hyperplane approximation allows a very precise real time tracking.

D. Template tracking.

Differential methods only allow small movement estimation. If, after a certain time, the reference template is far from its initial position, equations (3) and (4) can no longer be used to compute its movement. Thus the tracking stage needs a frame to frame modification of the reference template.

Let $\boldsymbol{\beta}_k$ be the vector representing the position of the template in image number k. In image number k+1, the position of the template will be referred to as $\boldsymbol{\beta}_{k+1}=\boldsymbol{\beta}_k+\boldsymbol{\partial\beta}$. Let $I_k(\mathbf{x})$ be the illumination at location $\mathbf{x}$ on image number k. If unexpected variations ε_i are neglected, then equation (1) becomes:

$$I_k(f(\mathbf{x_i},\boldsymbol{\beta}_k)) \approx I_{k+1}(f(\mathbf{x_i},\boldsymbol{\beta}_k+\boldsymbol{\partial\beta})) \approx I_0(\mathbf{x_i}) \quad (5)$$

Then differential equation (2) becomes:

$$I_{k+1}(f(\mathbf{x_i},\boldsymbol{\beta}_k+\boldsymbol{\partial\beta})) - I_k(f(\mathbf{x_i},\boldsymbol{\beta}_k)) \approx \mathbf{M}.\boldsymbol{\partial\beta} \quad (6)$$

Now $I_k(f(\mathbf{x_i},\boldsymbol{\beta}_k)) \approx I_0(\mathbf{x_i})$ therefore:

$$I_{k+1}(f(\mathbf{x_i},\boldsymbol{\beta}_k+\boldsymbol{\partial\beta})) - I_0(\mathbf{x_i}) \approx \mathbf{M}.\boldsymbol{\partial\beta} \quad (7)$$

Which means that the computation of $\boldsymbol{\partial\beta}$ only requires a comparison between the illumination of the reference image I_0 at location $\mathbf{x_i}$ with the illumination at location $f(\mathbf{x_i},\boldsymbol{\beta}_k)$. Then the position parameter $\boldsymbol{\beta}_{k+1}$ is obtained by:

$$\boldsymbol{\beta}_{k+1}=(\boldsymbol{\beta}_k+\boldsymbol{\partial\beta}) \quad (8)$$

E. Limitation due to discretization.

The relation between movement and grey level variation is modeled in a way that does not take the discrete and bounded nature of the used information into account. Estimation of the invert problem strongly relies on the hypotheses that errors can more or less be assumed to be random centered and symmetrical noise (use of least squares). Discretization is assumed to produce simple additive noise. The fact that grey level variation and image location are bounded is ignored.

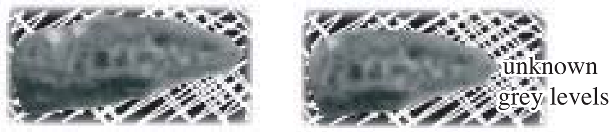
Ignoring the discrete nature of both spatial localization and grey level variation substantially limits the power of the proposed solutions. In fact, normal hypotheses of discretization effects can be assumed when the number of pixels of the reference template is large enough to apply the central limit theorem. The tracking is biased when there are small patterns.

Ignoring the bounded nature of both spatial localization and grey level variation limits the use of derivative relations to pixels that are far from the bounds with respect to localization (side effect) or grey level (saturation). Unfortunately, the most reliable pixels are those whose illumination is close to saturation. When using standard least square methods, such pixels tend to bias the global estimation. If robust estimation is used, saturated pixels are rejected as outliers.

F. Limitations due to modeling.

In this framework, global or local changes in illumination can bias the movement estimation. It is possible to take a change of illumination into account by increasing the size of parameter $\boldsymbol{\beta}$. However, the size of the template has to be also increased to avoid bias due to over-learning.

Finally, in both tracking and learning processes, it is necessary to be able to represent a total ignorance of information especially because of boundary effects. In the learning stage, for example, estimation of matrices $\mathbf{M}$ or $\mathbf{H}$ are computed by synthetically moving the reference frame. To prevent the learning stage to take information from the background into account, values outside of the template should be avoided (Fig. 2). This is not possible with classical learning processes.



reference template before ... and after a synthetic motion

Figure 2: grey levels are unknown outside the template

If, during the tracking stage, the template reaches the bounds of the image, then an analog situation will apply when information about the grey level local distribution is not available.

A new framework is required to overcome all these limitations. In this paper, we propose to shift from classical statistics to possibilistic bipolar logic.

3. Optical flow with bipolar logic.

A. Toward a new framework.

In a previous article [4], we presented an alternative to differential optical flow like computation of main apparent motion. In this new framework, the localization imprecision due to sampling was represented using a partition of the real plane $\mathbb{R}^2$ with fuzzy quantities. Grey level representation assumes that the photoreceptive sensors of the camera perform a fuzzy (imprecise) classification, splitting the set of image pixels into two dual classes: bright (white) pixels and dark (black) pixels. This classification is assumed to apply when the template is moving, which is a very weak assumption. The middle gray pixels are said to be least reliable. The algorithm derived from this framework is based on a Hough-like imprecise vote procedure using bipolar possibilistic logic. The use of a quasi-continuous histogram technique [5] allows accurate and robust real modal estimation of motion.

However, this method is not suitable for template tracking. As an area-based matching technique, its computing complexity is very high. The computing time can be quite long since voting and searching loops are intertwined. Sub-sampling techniques have to be used for high order parametric models

In the method presented here, the relation between movement and grey level variation involves bipolar fuzzy rules.

B. Associating motion and grey-level clusters with bipolar rules — a brief introduction.

Bipolar logic is very close to ternary logic. Instead of using three symbols (1=true, 0=false,?=unknown), an item of information A is associated with the pair $(\Pi(A), \Pi(A^c))$ where A^c is the contrary event.

To simplify the explanation and illustrations of this new concept, we will present the basis of grey-level-cluster-to-motion association with a simple binary rectangular template.

Pixels belonging to the template are numbered from 1 to 16, as shown in Fig. 3. Each pixel is associated with a binary grey level g_i. The pixels are split in two dual clusters (white pixels and black pixels) according to bipolar possibilistic logic. If the binary grey level value of the pixel located at $\mathbf{x}_i$ belongs to the white cluster (rsp. black cluster) $\prod_W(\mathbf{x}_i)=1$ and $\prod_B(\mathbf{x}_i)=0$ (rsp. $\prod_W(\mathbf{x}_i)=0$ and $\prod_B(\mathbf{x}_i)=1$). If the binary grey level value g_i of the pixel located at $\mathbf{x}_i$ is unknown $\prod_W(\mathbf{x}_i)=1$ and $\prod_B(\mathbf{x}_i)=1$. The template and dual clusters are presented in Fig. 3.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

binary template to be tracked

1	1	0	1
1	1	0	1
0	0	0	1
1	1	1	1

white pixels of the template

0	0	1	0
0	0	1	0
1	1	1	0
0	0	0	0

black pixels of the template

Figure 3: Binary template

Let us now assume that the template movement can be represented by a simple translation of one pixel in each direction. T_x is a horizontal translation towards the right and T_y is a vertical down translation. All possible motion is represented by a vector $\mathbf{a}_{s\ (s=1...9)}$.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

synthetically translated template

1	0	1	1
1	0	1	1
0	0	1	1
1	1	1	1

white pixels

0	1	0	1
0	1	0	1
1	1	0	1
0	0	0	1

black pixels

Figure 4: Translation of one pixel towards the left

Fig. 4 shows the reference template modified by the movement represented by $\mathbf{a}_4$, i.e. a translation of one pixel toward the left (T_x =-1, T_y=0). If the outside of the template is unknown, the grey-level of the pixels 4, 8, 12 and 16 (grey-colored) are unknown after the translation. They can belong to both clusters (black or white).

We propose to link each motion $\mathbf{a}_s$ to the grey level cluster associated to each pixel $\mathbf{x}_i$ of the template by using two dual rules: *white rule* $(\mathbb{R}_W^{i,s})$ and *black rule* $(\mathbb{R}_B^{i,s})$.

$\mathbb{R}_W^{i,s}$: *If the template is modified with the motion $\mathbf{a}_s$ then the pixel located at $\mathbf{x}_i$ is white.*

$\mathbb{R}_B^{i,s}$: *If the template is modified with the motion $\mathbf{a}_s$ then the pixel located at $\mathbf{x}_i$ is black.*

Clearly, possibility of the relation $\mathbb{R}_W^{i,s}$ is linked to the possibility that the pixel located at $\mathbf{x}_i$ is be white. While necessity of this rule is linked to the impossibility that this pixel is black. Since the two clusters are dual, all information about these rules can be represented by $\Pi(\mathbb{R}_W^{i,s})$ and $\Pi(\mathbb{R}_B^{i,s})$.

For example, concerning the pixel located at $\mathbf{x}_6$ of the reference template and the movement associated with $\mathbf{a}_4$, $\prod(\mathbb{R}_W^{6,4})=0$ and $\prod(\mathbb{R}_B^{6,4})=1$. Concerning the pixel located at $\mathbf{x}_8$ on the reference template, $\prod(\mathbb{R}_W^{8,4})=1$ and $\prod(\mathbb{R}_B^{8,4})=1$ because the grey level of this pixel is unknown after the movement associated with $\mathbf{a}_4$. Therefore it represents the lack of knowledge about the class to which the pixel located at $\mathbf{x}_8$ will belong after a movement $\mathbf{a}_4$.

C. Abduction-based rule inversion.

Abduction is a reasoning process that seeks possible explanations for abnormal observations. The bipolar rules we use allows us to deduce the grey level cluster of each pixel of the reference template from the movement knowledge. Abduction-based reasoning has to be used to determine the movement of the template from the new grey level assignment.

Let us explain abductive inversion concerning the pixel of the template whose location is $\mathbf{x}_6$. The rule $\mathbb{R}_B^{6,4}$ (if the movement is $\mathbf{a}_4$ then the pixel located at $\mathbf{x}_6$ is black) is true, i.e. $\prod(\mathbb{R}_W^{6,4})=0$ and $\prod(\mathbb{R}_B^{6,4})=1$. If, after a movement, pixel $\mathbf{x}_6$ is black, no information can be obtained on the projected movement. However, pixel $\mathbf{x}_6$ is white, then the movement $\mathbf{a}_4$ has to be discarded. This is abductive reasoning.

D. Fuzzy partition of grey-level.

To account for the grey level information, we propose to use a fuzzy partition of the grey-level values. Fig. 5 shows this fuzzy partition. $\mu_B(g)$ (rsp. $\mu_B(g)$) is the membership function of the grey level g to the white (rsp. black) cluster.

This partition allows a straighforward representation of the quantization. Let us suppose that the digital grey level of the pixel located at $\mathbf{x}_i$ is $g_i=I(\mathbf{x}_i)=100$. Imprecision due to sampling is associated with an interval whose width is 1. Then the grey level value is an interval [100, 101]:

$$\prod_W(\mathbf{x}_i) = \prod(\mathbf{x}_i \text{ is white}) = \operatorname*{\mathbf{Sup}}_{g \in [100,\, 101]}(\mu_W(g)) \approx 0{,}396$$

$$\prod_B(\mathbf{x}_i) = \prod(\mathbf{x}_i \text{ is black}) = \operatorname*{\mathbf{Sup}}_{g \in [100,\, 101]}(\mu_B(g)) \approx 0{,}608$$

$$N_W(\mathbf{x}_i) = 1 - \prod_B(\mathbf{x}_i) \approx 0{,}392 < \prod_W(\mathbf{x}_i).$$

If the grey level of the pixel located at $\mathbf{x}_i$ is unknown then:

$$\prod_W(\mathbf{x}_i) = \operatorname*{\mathbf{Sup}}_{g \in [0,\, 255]}(\mu_W(g)) = 1 \text{ and } \prod_B(\mathbf{x}_i) = \operatorname*{\mathbf{Sup}}_{g \in [0,\, 255]}(\mu_B(g)) = 1$$

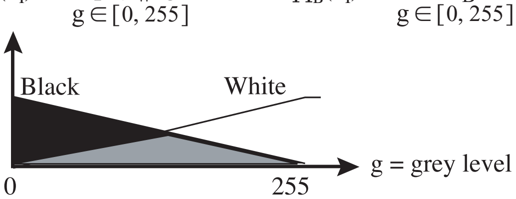


Figure 5: Linear membership function.

E. Fuzzy movement.

The parameter space representing the movement is partitioned in p fuzzy-subsets. Each fuzzy subset $\mathbf{A}_s$ is an imprecise domain that restricts parameter **ß** which describes the movement. Fig. 6 represents a regular fuzzy partition of p=25 clusters of the space of horizontal and vertical translations.

F. Fuzzy pixel.

We take the imprecision induced by sampling of the image into account by considering each pixel as a fuzzy quantity of $\mathbb{R}^2$. Each image pixel is represented by the fuzzy box $\mathbf{X}_i$ (Fig. 7) that restricts the possible locations $\mathbf{x}$ of the pixel whose grey level is $g_i=I(\mathbf{x_i})$. The mode of the fuzzy pixel is defined by $\mathbf{x}_i$. Its support is defined by its neighbor pixels.

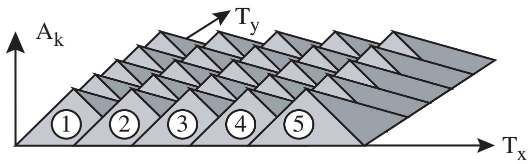


Figure 6: Partition of the translation space.

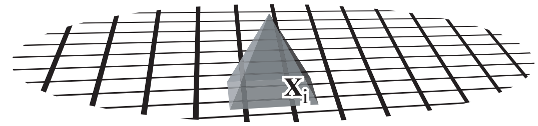


Figure 7: fuzzy pixel $\mathbf{X}_i$.

G. Fuzzy rules.

If both pixel and movement are represented by fuzzy subsets, then the computation of the possibilities $\prod(\mathbb{R}_W^{is})$ and $\prod(\mathbb{R}_B^{is})$ have to be modified. We use the extension principle [6] to compute the fuzzy domain $\mathbf{X}_i\text{'}=\mathcal{F}(\mathbf{X}_i,\mathbf{A}_s)$, i.e. the image of fuzzy pixel $\mathbf{X}_i$ by the function f parametrized by the fuzzy subset $\mathbf{A}_s$ (Fig. 8).

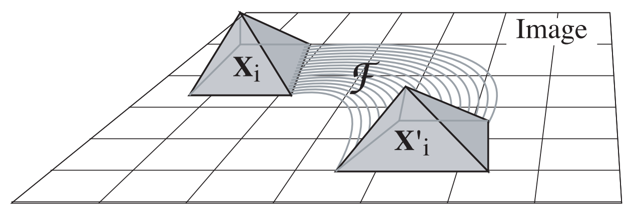


Figure 8: fuzzy movement.

Computation of the possibility of the two dual rules associated with $\mathbf{X}_i$ and $\mathbf{A}_s$ involves computation of the possibility that the domain $\mathcal{F}(\mathbf{X}_i,\mathbf{A}_s)$ belong to the white (rsp. black) cluster. Let $\wedge$ be the T-norm, this computation is given by:

$$\Pi(\mathbb{R}_W^{i,s}) = \operatorname*{Sup}_k \left\{ \Pi(\mathcal{F}(\mathbf{X}_i, \mathbf{A}_s); \mathbf{X}_k) \wedge \Pi_W(\mathbf{X}_k) \right\} \quad (9)$$

$$\Pi(\mathbb{R}_B^{i,s}) = \operatorname*{Sup}_k \left\{ \Pi(\mathcal{F}(\mathbf{X}_i, \mathbf{A}_s); \mathbf{X}_k) \wedge \Pi_B(\mathbf{X}_k) \right\} \quad (10)$$

H. Fuzzy abduction.

Here we estimate the *vote* of the pixel number i of the template for the movement $\mathbf{A}_s$. This vote is imprecise because the information that enables this vote is imprecise. The upper and lower bounds of the vote are given by $\prod(\mathbf{A}_s;\mathbf{X}_i)$ and $N(\mathbf{A}_s;\mathbf{X}_i)$ computed by abduction-based reasoning:

$$N(\mathbf{A}^c{}_s;\mathbf{X}_i) = \bigvee \begin{pmatrix} [N(\mathbb{R}_W^{i,s}) \wedge (1 - \Pi_W(\mathbf{X}_i))] \\ [N(\mathbb{R}_B^{i,s}) \wedge (1 - \Pi_B(\mathbf{X}_i))] \end{pmatrix} \quad (11)$$

$$\prod(\mathbf{A}^c{}_s;\mathbf{X}_i) = \bigvee \begin{pmatrix} [\Pi(\mathbb{R}_W^{i,s}) \wedge (1 - N_W(\mathbf{X}_i))] \\ [\Pi(\mathbb{R}_B^{i,s}) \wedge (1 - N_B(\mathbf{X}_i))] \end{pmatrix} \quad (12)$$

$\prod(\mathbf{A}_s;\mathbf{X}_i)= 1-N(\mathbf{A}^c{}_s;\mathbf{X}_i)$ and $N(\mathbf{A}_s;\mathbf{X}_i)= 1-\prod(\mathbf{A}^c{}_s;\mathbf{X}_i)$.

with $\prod_{W/B}(\mathbf{X}_i)=\prod(\text{pixel located at } \mathbf{X}_i \text{ is white / black})$. The pair $[N(\mathbf{A}_s;\mathbf{X}_i), \prod(\mathbf{A}_s;\mathbf{X}_i)]$ is the imprecise vote of the pixel approximately located at $\mathbf{X}_i$ to the imprecise domain $\mathbf{A}_s$ of the movement parameter $\boldsymbol{\beta}$.

4. Movement estimation: use of quasi-continuous histograms.

The previous stage gives a vote of each pixel imprecisely located at $\mathbf{X}_i$ for the movement represented by the imprecise parameter $\mathbf{A}_s$. Known imprecisions due to modeling of all information involved in the process are taken into account and propagated to the imprecise vote (the greater the information imprecision, the greater the vote imprecision). $N(\mathbf{A}_s;\mathbf{X}_i)$ is called the pro-vote (or necessary vote) and $\prod(\mathbf{A}_s;\mathbf{X}_i)$ is called the supporting vote (or possible vote).

All variations that affect the link between the grey-level distribution and pattern motion are not represented in this modeling. Thus, a conjunctive fusion of the votes would result in a void decision. Therefore we use a statistical polling process via quasi-continuous histograms.

A. Quasi-continuous histograms (QCH).

The theoretical framework of quasi-continuous histogram (QCH) is based on the fuzzy rough sets theory [7]. The main goal is to provide a framework that generalizes the histogram concept by dissociating the histogram's granularity from the precision of the information computed using this histogram [5]. QCH are built on a fuzzy partition to lower the effect of the arbitrary partitioning. QCH accounts for the imprecision and uncertainty of the data to perform moment, modal or rank statistics.

Here we are interested in modal statistics, i.e. finding the movement which maximizes the number of votes [4]. In this case, an imprecise accumulator is associated to each cluster $\mathbf{A}_s$ of the partition of the motion parameter space. An imprecise accumulator is an interval whose lower bound is the sum of the lower votes (necessary votes) and the upper bound is the sum of upper votes (possible votes):

Searching the main mode in a QCH amounts to searching for the position of a crisp (or fuzzy) subset **B** with a certain granularity Γ polling a locally or globally maximum number of votes [4]. Usually **B** is chosen to be a crisp or fuzzy symmetric quantity. To perform this maximization, the number of votes purporting to any subset **B** has to be estimated. This estimation is achieved by transferring the imprecise number of votes of each cluster A_s to the subset **B**. The pignistic transfer we use is defined in [9].

This technique gives a *robust* estimation of **B** in that sense that the estimation is insensitive to small deviations from the assumptions. Phenomena like occlusion, change of reflective properties, or orientation variations have little effect on the estimation since few pixels are altered. The insensitivity to illumination changes is due to the grey-level pre-classification.

Finally, the value of $\partial\boldsymbol{\beta}$ assigned to the movement of the template between consecutive images is given by the mode, or the center of gravity of the subset **B**.

B. Tracking and update.

Let us suppose that the position $\boldsymbol{\beta}_k$ of the template on image k is known. The position of each fuzzy pixel $\mathbf{X}_i'$ of the template can thus be computed by using the extension principle: $\mathbf{X}_i'=\mathcal{F}(\mathbf{X}_i,\boldsymbol{\beta}_k)$. For each pixel $\mathbf{X}_i'$ the possibilities $\prod_W(\mathbf{X}_i')$ and $\prod_B(\mathbf{X}_i')$ are computed on the $(k+1)^{th}$ image using the Sup-min principle. Abduction based reasoning provides an imprecise estimation of the vote of each pixel $\mathbf{X}_i'$ for each fuzzy movement $\mathbf{A}_s$. These imprecise votes are accumulated in the QCH associated with the partition of the parameter space. The estimation step provides the restriction **B** of the differential position of the template. The mode of **B** is used as the estimation of $\partial\boldsymbol{\beta}_{k+1}$. Finally, $\boldsymbol{\beta}_{k+1}$ is given by updating $\boldsymbol{\beta}_k$ with $\partial\boldsymbol{\beta}_{k+1}$.

5. Experimentations.

We performed numerous experiments on real and synthetic image sequences in order to test the ability of the tracker to follow the template in various conditions. We present experiments that highlighted some noteworthy properties of the tracker. All experiments presented here are based on a planar translation motion model.

A. Precision

Figure 9: tracking Lena's eye.

To evaluate the precision of the estimator, we generated various image sequences from a single image. We illustrate these experimentations with results obtained by randomly translating the famous *Lena* image (Fig. 11). The movement is planar with maximum amplitude of 15 pixels. The tracking template is made of about 1500 pixels around the left eye. For sake of simplicity, the discretization step of the parameter space (translation space) is 1 pixel. The synthetic images are obtained by a bicubic interpolation.

The granularity of the histogram (surface of each cluster $\mathbf{A}_s$) is equal to 1. The error is centered (mean equals 0), the average error is close to 0.07 and the maximum error is less than 0.3. Note that the error is lower than the histogram granularity.

B. Robustness.

The robustness of the estimator was tested on a standard PC pentium IVM 2.2 GHz workstation with a standard Web Cam. The templates were manually selected. In these experiments, the camera and template were moving with motions that corresponded or not to the modeled motion. We illustrate these experiments with two sequences of among 200 frames involving a toy-robot. We tried to track the head of the robot. In the first sequence, the camera is rotating around the vertical and the horizontal axes. The robot is not moving (Fig. 10). In the second sequence, the robot is moving towards the camera (Fig. 11).

Figure 10: Still robot with moving camera.

Figure 11: Moving robot with still camera.

In both cases, the projected movement does not fulfill the working hypothesis. The motion is assumed to be a simple translation while the real movement involves zoom effects and rotations. The tracker does not drop out while the movement can still be approximated by a simple translation, i.e. the hypothesis is fulfilled by around 50% of the template.

The last group of experimentations revealed the insensitivity of this method to illumination variation. To illustrate this ability, we synthetically increased the brightness of the lizard sequence (16 frames) up to 30% of the full grey level range. Fig 12 shows the first and last image of the sequence.

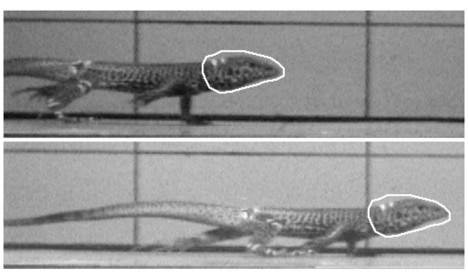


Figure 12: Brightness variation.

6. Conclusion and discussion.

We have presented an original framework for performing template tracking in a video sequence. This method is based on a new representation and inversion of the link between grey-level variation and projected motion. This representation involves bipolar partitioning of the grey-level space, assuming that the photo-receptive sensors of the camera perform a fuzzy (imprecise) classification of the image. It is assumed that there will be compliance to classification while the template is moving. The projected motion of the template is assumed to be described by a parameterized function. The parameter space is partitioned in a fuzzy way. We use the extension principle to estimate the possibility of linking each cluster of the parameter space to the grey level classification of each pixel of the template. This estimation is performed in the first image of the sequence.

While the object is moving, we use abduction-based reasoning to deduce, from the rules and the new grey level assignment of the pixels of the template, the vote of each pixel to each motion cluster. Then we use the quasi-continuous histogram technique to find the *best* parameter fuzzy subset i.e. the subset which polls a locally or globally maximum number of votes. Then the position of the template is updated.

Our method allows real-time tracking on a standard workstation (less than 1/25 sec.) in case of simple movements. Numerous experimentations have shown that this solution is efficient and robust with respect to outliers and unexpected phenomena like change of illumination, partial occlusions, pose variation and movements that are not taken into account in the model (in this paper zoom effect and rotations). This robustness is obtained without any modification of the algorithm while other methods require explicit modeling of unexpected variations like time-varying illumination.

Usually, with differential methods frame to frame changes have to be small, limiting the speed at which the template can move. Our algorithm does not have the same kind of limitation. Its ability to perform movement estimation is linked to the coverage of the parameter space by the fuzzy partition. It is also linked to the intersection area of the target region before and after movement. If this area is less than 50% of the template area, the estimation cannot be guaranteed. To improve tracking, a straightforward solution is to use a multi-hypothesis algorithm based on a (fuzzy) set description of the parameter.

In the future, we plan to perform an objective comparison of the area-based and differential-based tracking methods involving a benchmark of image sequences with known realistic movement. We are also investigating the application of this method to perform template recognition. Finally, we study a simplification of the algorithm to allow a high order representation of movement using a coarse to fine strategy.

7. References

[1] G. Hager, P. Belhumeur. *Efficient region tracking with parametric models of geometry and illumination*. IEEE Transactions on P.A.M.I., 20(10), pp. 1025-1039, 1998.

[2] F. Jurie, M. Dhome. *Hyperplane approximation for template matching*. IEEE Transactions on P.A.M.I., 24(7), pp. 996-1000, 2002.

[3] M. La Cascia, S Sclaroff. *Fast, reliable head tracking under varying illumination: an approach based on registration of texture-mapped 30 models*. IEEE Transactions on PAMI, Vol. 22, No. 4, april 2000, pp. 322-336.

[4] F. Comby, O. Strauss, M.J. Aldon, *Possibility theory and rough histograms for motion estimation in a video sequence*, IWVF'O1: 4th International Workshop on Visual Form, Capri, Italy, May 28-30 2001.

[5] O. Strauss, F. Comby and M.J. Aldon. *Rough histograms for robust statistics*. Proceedings of the International Conference on Pattern Recognition, vol2, pp. 688-691. Sept. 2000.

[6] D. Dubois, H. Prade. *Possibility theory, an approach to computerized processing of uncertainty*, Plenum Press, New York 1988.

[7] D. Dubois, H. Prade. *Rough fuzzy sets and fuzzy rough sets*. International journal of general systems, 17(2-3), pp191-200, 1990.

[9] Smets Ph. and Kennes R., *The transferable belief model*, Artificial Intelligence, 1994, 66:191-234.