# Background Subtraction under Uncertainty using a Type-2 Fuzzy Set Gaussian Mixture Model

Ali DARWICH*† , Pierre-Alexandre HÉBERT*, Yasser MOHANNA† and André BIGAND *
*University of Littoral Côte d'Opale (ULCO), Calais, France
†Lebanese University
Email: hebert@univ-littoral.fr, bigand@univ-littoral.fr

***Abstract*—Detection of moving objects is an essential step of vision processing commonly used in video surveillance systems. Algorithms based upon adaptable background modelling are often chosen to provide a robust object detection, even when the background is changing. The proposed approach exploits a type-2 fuzzy set Gaussian Mixture Model, whose main interest lies in its ability to take robust decisions in spite of uncertain data and model. Object detection results are supplied by fuzzy values, whose modelled uncertainty can benefit - as shown - to a post-processing. The proposed post-processing consists in a spatial fusion of the detection responses in the neighborhood of pixels. The presented experimental results show the efficiency of this original method.**

***Index Terms*—Type-2 Fuzzy sets; Video Background Subtraction; Gaussian Mixture Model.**

## I. Introduction

Object detection is a crucial preparatory step of many computer-vision applications. A common approach consists in a background subtraction, which classifies as "foreground", pixels of a video frame that differ enough from pixels of an estimated background frame. Detection of moving objects - by subtracting an estimated background image to each frame - may appear as an easy task. But many issues prevent such a basic technique to get good background subtraction results. Indeed, it often fails in real videos: for instance when the image is spoiled by noise, or when a little change in the environment disturbs the motion detection - typically illumination, shadow, slight movement and so on. For this reason, a lot of research has been devoted to develop new techniques to overcome the main obstacles.

We focus here on pixel-level background modellings, which uses a probabilistic density function to model the frequency of pixel values over the video frames. Decision is then frequently obtained from a *maximum likelihood*-like estimation method. Many algorithms using Gaussian models have been proposed [1]–[9]. Most of them use a single Gaussian density function per pixel, to model its intensity distribution. However, pixel values may have more complex distributions and Gaussian mixtures are generally preferred.

Stauffer developed early one of the most important Gaussian Mixture Models (GMMs)-based algorithms for real-time background subtraction [7]. He proposed first a multi-modal distribution allowing a complex background modelling, then an algorithm to update this model in real-time. Each Gaussian mode is then assumed to be modelling either background pixel values (the most frequent values) or foreground pixel values (the less frequent ones). This model is able to fit background changes, more or less quickly according to an adjustable learning rate.

Bouwmans *et al.* provided a review and an original classification of the numerous improvements of this initial GMM subtraction algorithm [10].

More recently, Chen *et al.* proposed a background subtraction algorithm based on a GMM modelling and on post-processings used to improve the decisions: minimum spanning tree (a spatial fusion step) and optical flow estimation using robust estimators M-smoother (a temporal fusion step) [11].

An important issue of those GMM methods relies in the ambiguity of the "most likely mode" (either foreground or background) associated to a given pixel value. This happens when two Gaussian modes are very close and overlap. Some fuzzy methods were introduced to take into account this ambiguity and the imprecision of mode distributions. In particular, Bouwmans *et al.* proposed a background modelling method [8] based on type-2 fuzzy GMMs [12]. But as further explained, the introduced fuzziness is not fully exploited, because of a crisp mode selection.

In this paper, we propose to take into account the ambiguity of the mode selection, with a type-2 fuzzy approach. Unlike most methods in literature, the mode selection and the intermediate decision are no more crisp but fuzzy. And this uncertainty is used in a final spatial fusion step, to average the fuzzy decision in the neighborhood of pixels. After this fusion, a final binary decision is processed. This uncertainty handling aims at improving the robustness of the detection of the dynamic background.

The paper is organized as follows: Section 2 introduces Gaussian Mixture Models in background modelling, and its type-2 fuzzy extension. Section 3 proposes a background subtraction under uncertainty algorithm using a type-2 fuzzy set Gaussian Mixture Model. And Section 4 presents some results on two real videos, whose one comes from the *Change Detection 2014 benchmark dataset* [13].

## II. Toward Fuzziness in Gaussian Mixture Models for Background Subtraction

### A. GMM for Background Subtraction

Distribution of data coming from several groups may be well modeled by a probabilistic mixture model, with one distribution component (or mode) per group. In the current application, values of a pixel over time may be characterized by such a mixture model: each mode being either associated to the background or to the foreground.

Gaussian Mixing Model (GMM) is the most popular technique to model the [8] background and foreground state of a pixel.

Let $I_t$ be the video frame number $t$ and $p$ the studied pixel - of coordinates $(i, j)$ - and $x_t^p$ be its value in frame $I_t$. The sample of values of this particular pixel over time is then denoted as:

$$\{x_1^p, \cdots, x_t^p\} = \{I_t(i, j) : 1 \le t \le T\}, \tag{1}$$

with $T$ the number of frames.

The GMM associated to pixel $p$ at frame $t$ is composed of $K$ weighted Gaussian functions. For the sake of simplicity, those normal distributions are choosen *univariate* in the whole paper:

$$P_t^p(x) = \sum_{i=1}^{K} w_{i,t}^p . f(x; m_{i,t}^p, \sigma_{i,t}^p), \tag{2}$$

with:

- $K$: the number of mixture components,
- $f(x; m_{i,t}^p, \sigma_{i,t}^p)$: the Gaussian density function of the $i-$th component of pixel $p$'s GMM at frame $t$,
- $w_{i,t}^p$: its weight,
- $m_{i,t}^p$ and $\sigma_{i,t}^p$: its parameters (mean and variance):

$$f(x; m_{i,t}^p, \sigma_{i,t}^p) = \frac{1}{\sigma_{i,t}^p \sqrt{2\pi}} \exp^{\frac{-(x-m_{i,t}^p)^2}{2\sigma_{i,t}^{p\,2}}}. \tag{3}$$

Figure 1 shows a particular pixel $p_0$ of coordinates $(i_0, j_0)$ in two different frames. Because of a moving character, its state changes from foreground to background.

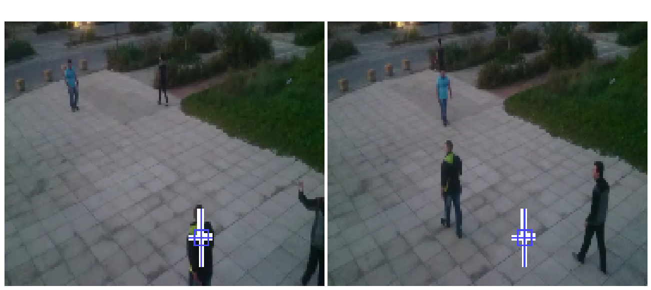

Figure 1. Two different frames (1 and 100) with marking of the studied pixel.

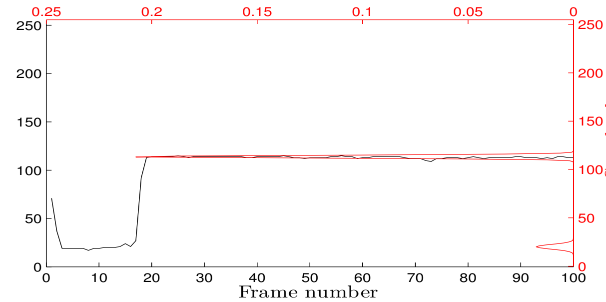


Figure 2. Pixel intensity evolution.

Figure 2 shows the evolution of this pixel over time (in black), and its final GMM $P_T^{p_0}(x)$ (in red). The important gap occurring in the first frames is due to the moving character. It results in a *low weighted* Gaussian distribution, which should normally be interpreted as a foreground mode.

### *B. Type-2 Fuzzy Sets*

*1) General framework:* Type-2 Fuzzy sets (T2 FSs) are often used to model the uncertainties in real applications: imprecision, noise, incompleteness. Unfortunately, T2 fuzzy sets are more difficult to use and understand than ordinary (ie. type-1) fuzzy sets, therefore their use is not yet widespread, especially in image processing.

A fuzzy set $\tilde{A}$ allows to model the imprecision of a variable $A$: either a quantitative variable, whose measurements are not precisely known; or a qualitative variable, such as a linguistic one [14], whose mapping to a quantitative domain is not precisely defined.

In contrast to crisp sets, the membership of element $x \in X$ to a fuzzy set is not binary. A fuzzy set $\tilde{A}$ is defined by its *membership function* (MF) $\mu_{\tilde{A}}$, which maps each element of domain $X$ to a membership degree.

In T1 FSs the membership degrees are numbers in $[0, 1]$, whereas in T2 FSs the membership degrees are themselves fuzzy sets [15] (called a "secondary membership function"). The T2 FS was introduced by Zadeh in 1975 and was extended by Karnik and Mendel in 1998 [16], [17], and Turksen in [18] for ontologies.

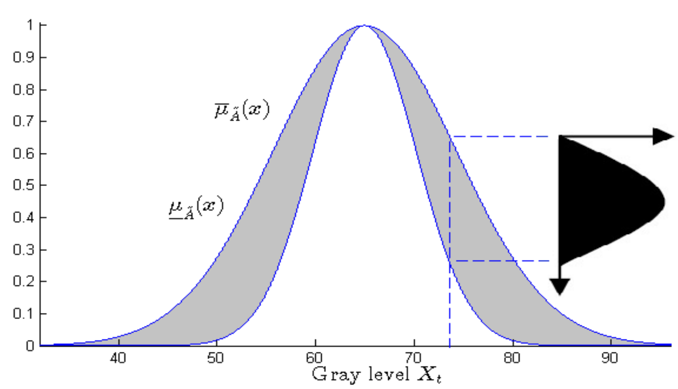


Figure 3. A type-2 fuzzy set.

In Figure 3 the shaded region is the footprint of uncertainty (FOU) defined by the secondary membership function (which appears in black on the right-hand side of the the figure). The upper bound function corresponding to the maximum membership grade of the FOU is the upper MF (defined as $\overline{\mu}_{\tilde{A}}$ in the sequel), and the bound that has the minimum membership grade of the FOU is the lower MF (defined as $\underline{\mu}_{\tilde{A}}$). Let $X$ be the universe of discourse, the membership grade of a T2 FS $\tilde{A}$, associated to a particular value $x$, is defined by the following mapping:

$$\mu_{\tilde{A}} : X \to [0,1]^{[0,1]}. \tag{4}$$

Thus the T2 FS $\tilde{A}$ can be defined as:

$$\tilde{A} = \{(x,u), \mu_{\tilde{A}}(x,u) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\}, \tag{5}$$

with $J_x$ the support of the secondary MF of $\tilde{A}$ in $x$.

*2) Interval-Valued Fuzzy Sets:* General T2 FS framework may be simplified, by defining secondary membership functions as crisp intervals: such T2 FS are called *Interval-Valued Fuzzy Sets* (IVFSs) [19]. Two membership functions are enough to model an IVFS:

- T2 FS upper membership function: $\overline{f}(x)$.
- T2 FS lower membership function: $\underline{f}(x)$.

This simplification of T2 FSs represents a crucial advantage, from a software implementation point of view. This is why IVFSs are the most used T2 FSs in image processing [20].

Those two membership functions are frequently built from a single ordinary Gaussian function. A simple procedure consists in introducing imprecision in its mean parameter [21]. $m$ is no more precisely known, but known as bounded by a $\Delta_m$-half-length interval centered in $m_c$:

$$m \in \left[\underline{m} = m_c - \Delta_m, \overline{m} = m + \Delta_m\right]. \tag{6}$$

Lower and upper membership functions are then defined by shifting the mean inside its domain, and retaining maximal and minimal values of the Gaussian function $f$:

$$\overline{f}(x; m_c, \sigma, \Delta_m) = \underset{m \in [m_c - \Delta_m, m_c + \Delta_m]}{max} f(x; m, \sigma), \tag{7}$$

$$\underline{f}(x; m_c, \sigma, \Delta_m) = \underset{m \in [m_c - \Delta_m, m_c + \Delta_m]}{min} f(x; m, \sigma). \tag{8}$$

Figure 4 illustrates this building. Let us note that the larger the mean interval, the larger the FOU (and hence the uncertainty) of the resulting T2 FS.

Upper and lower membership functions could be defined - in a way quite similar - by considering an imprecise variance [12]. In this paper, we consider only the previous procedure.

### C. Background modelling using T2-FGMM with uncertain mean

Bouwmans *et al.* introduced T2 Fuzzy sets to handle uncertainty in GMM mixtures used for background subtraction [8]. They applied the GMM uncertain model proposed by Zeng *et al.* in paper *T2 Fuzzy Gaussian Mixture Model* [12]. In fact,

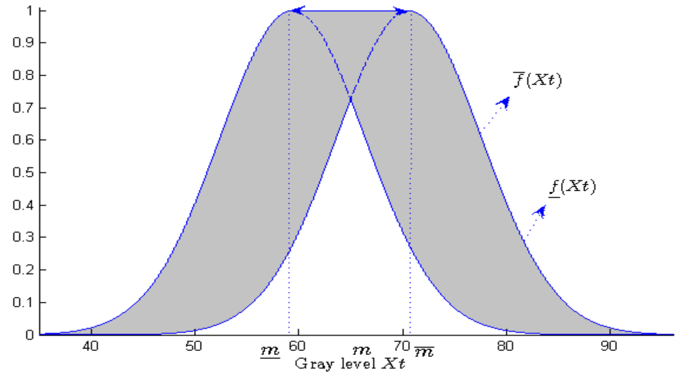


Figure 4. Gaussian interval type-2 fuzzy set (uncertain mean).

T2-fuzziness appears only in the final decision step of the method. Previous steps consist in a classical GMM approach.

The main steps of Bouwmans' algorithm are the followings:

For each pixel:

- ***Step 1*: GMM initialization:** a GMM is built (either trained from a set of values of the pixel, for instance from an EM algorithm, or initialized randomly).

For each frame:

- ***Step 2*: Modes labelling:** each Gaussian mode is labelled either *background* or *foreground*. This critical association is obtained from an empirical rule: the more frequent and precise the mode, the more likely it models *background* pixel-values.
  Concrtement, les modes $K$ sont tris selon leur indicateur $\frac{w_k}{\sigma_k}$. Next, the $K_B$ first modes are labelled *background*, thanks to a threshold $p \in [0,1]$:

$$K_B = \arg\min\left(\sum_{k=1}^{K_B} w_{k,t} > p\right). \tag{9}$$

  Other modes are labelled *foreground*.
- ***Step 3*: Pixel labelling:** the pixel is labelled according to the first mode compatible with its value (gray-level in our case). T2 FS are introduced in this mode selection (as explained in the next sub-section).
  In case no compatible mode is found, the pixel is set as *foreground*, and a new mode is created.
- ***Step 4*: GMM update:** if it already exists, the selected mode is reinforced (its weight $w_k$ increases); in the other case, it replaced the last sorted mode. Other modes are weakened.

The most important features of this algorithm are highlighted in the following sub-sections.

*1) T2 FS building for pixel labelling:* As emphasized, T2 FS appear in Step 3, when selecting the *likely* mode which best suits pixel value $x_t$. IVFS are built in the following way, for each Gaussian mode $k \in \{1, \ldots, K\}$:

- The $k$-th Gaussian mode function $f(x; m_k, \sigma_k)$ is first normalized (height set to 1) by the appropriate factor (this means that factor $\frac{1}{\sigma_k\sqrt{2\pi}}$ is ignored, as a simplification [12]).
- An IVFS is built from this function, by shifting the mean or the variance as previously explained. Without any loss of generality, we focus on the mean shifting.
  The quantity of uncertainty is set though a factor $k_m \geq 0$, tuning the length of the interval $[\underline{m}, \overline{m}]$ bounding the Gaussian mean parameter: $\Delta_{mk} = k_m * \sigma_k$.
  The obtained lower and upper membership functions are denoted $\underline{f}'_k$ and $\overline{f}'_k$.

*2) Pixel labelling from T2 FS:* Pixel $p$ is labelled with the label of the "best matching" Gaussian mode. This is the first ranked mode (according to $\frac{w_k}{\sigma_k}$) whose membership bounds in $x_t$ make this inequality true:

$$H_k(x_t) = \left| \ln\left(\overline{f}'_k(x_t)\right) - \ln\left(\underline{f}'_k(x_t)\right) \right| < k_t \sigma_k, \tag{10}$$

with $k_t$ a constant factor.

If this inequality cannot be asserted for any mode in $x_t$, then the last sorted mode is replaced by a new one, centered in $x_t$.

This choice is clearly justified by Bouwmans *et al.* They show that indicator $H_k(x_t)$ is a decreasing monotonous function of distance $|x_t - m_k|$:

$$H_k(x_t) = \begin{cases} 2\frac{k_m|x_t - m_k|}{\sigma_k}, \text{ if } x_t < \underline{m_k} \text{ or } x_t > \overline{m_k} \\ \frac{|x_t - m_k|^2}{2\sigma_k^2} + \frac{k_m|x_t - m_k|}{\sigma_k} + k_m^2, \text{ otherwise.} \end{cases}$$

So, the larger the normalized (log-likelihood) interval, the closer the pixel-value to the center of the mode. Consequently, the larger the interval, the more *likely* the mode.

Let us not that, because of their prior sorting, modes with highest weights and lowest variance - likely *background* modes - are tested first: if a pixel may belong to the background, it certainly does.

*3) Update of the GMM parameters:* After the mode selection step, the GMM parameters are updated, according to the following rule [7], [8].

If the selection succeeds, then the matching mode $i$ is "reinforced", and rejected modes $j \neq i$ are weakened:

$$w_{i,t+1} = (1-\alpha)w_{i,t} + \alpha \tag{11}$$
$$m_{i,t+1} = (1-\rho)m_{i,t} + \rho.x_{t+1} \tag{12}$$
$$\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + \rho(x_{t+1} - m_{i,t+1})^2 \tag{13}$$
$$w_{j,t+1} = (1-\alpha)w_{j,t}, \forall j \neq i, \tag{14}$$

with $\alpha$ a constant learning rate and $\rho = \alpha.f(x_{t+1}, m_i, \sigma_i)$.

Otherwise, the last distribution is replaced by a new Gaussian mode with the current value $x_t$ as its mean, an initially high variance, and a low weight parameter [7].

## III. Decision under uncertainty method using interval comparison

The algorithm proposed in this paper mainly differs from this previous work through two points:

- T2 FS uncertainty is used to handle the possible ambiguity of the mode estimation. This estimation - as in fuzzy classification, for instance in c-means algorithm - is no more binary, but truly fuzzy. Therefore, pixel may be partially labelled as both *background* and *foreground*.
- A spatial fusion step is added, in order to exploit this uncertainty. It is here considered that labels of pixels in a close neighborhood are presumably the same.

The proposed algorithm is illustrated in Figure 5.

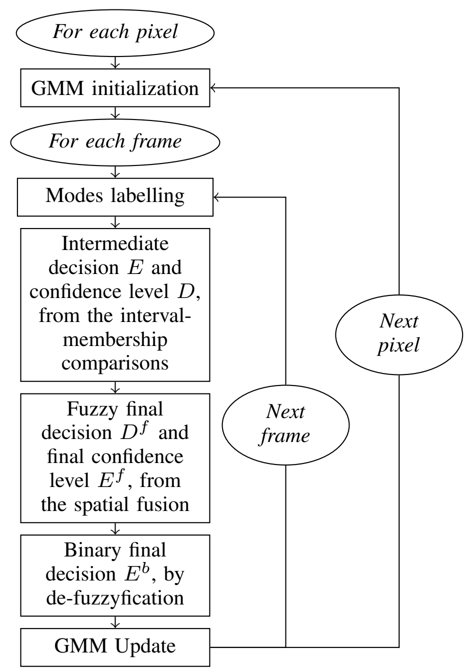


Figure 5. **Algorithm proposed.**

### A. Intermediate decision per pixel

Two first steps - **GMM initialization** and **Mode labelling** - are identical to Zeng's method [12].

First modification lies in **Pixel labelling** step. For each mode $k$, the IVFS associates to pixel $p$ in frame $t$ - of value $x_t$ - a normalized density interval: $[\underline{f}_k(x_t), \overline{f}_k(x_t)]$.

All those intervals are aggregated by a weighted average, according to their label. Weigths are defined by the values used by Zeng to sort the modes by "priority":

$$w'_k = \rho\left(\frac{w_k}{\sigma_k}\right), \tag{15}$$

with here $\rho(x) = x^2$.

More precisely, two label intervals are built by aggregation: $\mathbf{F}_t(p)$ to assess the hypothesis "pixel value matches a *foreground* region", and $\mathbf{B}_t(p)$ to assess the contrary hypothesis. A supplementary interval - called $\mathbf{R}$ for *reject* - is built to represent non-significant likelihood levels, from an a priori threshold $k_r \in \mathbb{R}^+$:

$$\mathbf{F}_t(p) = \frac{1}{\sum_{k=1}^{K} w'_k} \Big[ \sum_{k=K_B+1}^{K} w'_k \underline{f}_k(x_t), \sum_{k=K_B+1}^{K} w'_k \overline{f}_k(x_t) \Big] \quad (16)$$

$$\mathbf{B}_t(p) = \frac{1}{\sum_{k=1}^{K} w'_k} \Big[ \sum_{k=1}^{K_B} w'_k \underline{f}_k(x_t), \sum_{k=1}^{K_B} w'_k \overline{f}_k(x_t) \Big] \quad (17)$$

$$\mathbf{R} = [0, k_r]\,. \quad (18)$$

Then the fuzzy decision between *foreground* and *background* alternatives is built by comparing both intervals $\mathbf{F}_t(p)$ and $\mathbf{B}_t(p)$.

Comparison is achieved through a variant of the *acceptability index* proposed by Sengupta and Pal [22]. This index measures the "grade of acceptability of the first interval to be inferior to the second interval":

$$\mathscr{A}_{\otimes}(\mathbf{A}, \mathbf{B}) = \frac{c(\mathbf{B}) - c(\mathbf{A})}{\delta(\mathbf{B}) + \delta(\mathbf{A})}, \text{ if } c(\mathbf{A}) \leq c(\mathbf{B}) \quad (19)$$

with:

- $c(\mathbf{A})$ denotes the central value of interval $\mathbf{A}$;
- $\delta(\mathbf{A})$ denotes the half-width of $\mathbf{A}$.

We prefer to use a variant bounded in $[0, 1]$:

$$\mathscr{A}'_{\otimes}(\mathbf{A}, \mathbf{B}) = \begin{cases} \mathscr{A}_{\otimes}(\mathbf{A}, \mathbf{B}), & \text{if } c(\mathbf{A}) \leq c(\mathbf{B}) \\ & \text{and } \mathscr{A}_{\otimes}(\mathbf{A}, \mathbf{B}) \leq 1 \\ 0, & \text{if } c(\mathbf{A}) > c(\mathbf{B}) \\ 1, & \text{otherwise.} \end{cases} \quad (20)$$

Intermediate decision of pixel $p$ is finally built as follows:
$D_t(p) =$

$$\begin{cases} \frac{1}{2}\big[1 + \mathscr{A}'_{\otimes}\big(\mathbf{B}_t(p), \mathbf{F}_t(p)\big)\big], & \text{if } c\big(\mathbf{B}_t(p)\big) < c\big(\mathbf{F}_t(p)\big) \\ \frac{1}{2}\big[1 - \mathscr{A}'_{\otimes}\big(\mathbf{F}_t(p), \mathbf{B}_t(p)\big)\big)\big], & \text{if } c\big(\mathbf{F}_t(p)\big) < c\big(\mathbf{B}_t(p)\big) \\ \frac{1}{2}, & \text{if } c\big(\mathbf{F}_t(p)\big) = c\big(\mathbf{B}_t(p)\big). \end{cases} \quad (21)$$

So values greater (respectively lower) than 0.5 attest a likely *foreground* (r. *background*) decision, whereas values close to 0.5 just attest a lack of preference.

Finally, a confidence level $E_t(p)$ is computed to assess the level of significance of decision $D_t(p)$. The underlying idea is to measure how far the highest interval is from the reject interval $\mathbf{R}$:

$$E_t(p) = \max\{\mathscr{A}'_{\otimes}\big(\mathbf{R}, \mathbf{F}_t(p)\big), \mathscr{A}'_{\otimes}\big(\mathbf{R}, \mathbf{B}_t(p)\big)\}. \quad (22)$$

### *B. Final spatial fusion*

Each pixel $p$ in frame $t$ is now characterized by a fuzzy decision $D_t(p) \in [0, 1]$ and a confidence level $E_t(p) \in [0, 1]$. The decision value indicates which label *foreground* and *background* is preferred for this pixel. And the confidence value measures the quantity of information used in the decision computation. But finally a binary decision is expected.

This final step is added to show how uncertainty may be used to build a more robust final crisp decision.

The proposed procedure consists in a spatial fusion: intermediate fuzzy decisions of neighbor pixels are merged to try to avoid false detection on singleton pixels.

First a mask $3 \times 3$ (or larger) is centered on the current pixel $p$ and filled from its similarities with its 8 nearest neighbors. The similarity between pixel $p$ and its neighbor $p'$ is defined as follows:

$$S(p, p') = e^{-\left(\frac{x_t - x'_t}{\sigma_S}\right)^2}, \quad (23)$$

with $\sigma_S$ a tuning parameter.

Then mask values are weighted by the matching confidence degrees $E_t(p')$ of all neighbors $p'$ of $p$.

Next, this mask is multiplied element-by-element to a same-sized matrix, composed of the intermediate fuzzy decisions of the neighborhood of $p$; then its values are summed. So, the aggregated fuzzy decision of $p$ is defined as a weighted average:

$$D_t^f(p) = \frac{1}{\sum_{p' \in \mathcal{N}(p)} e_{p,p'}} \sum_{p' \in \mathcal{N}(p)} e_{p,p'}.D_t(p), \quad (24)$$

with:

- $\mathcal{N}(p)$: the 3x3 neighborhood of $p$ (including itself);
- $e_{p,p'} = S(p, p').E_t(p')$: the influence degree of pixel $p'$ on pixel $p$.

Final confidence level $E_t^f(p)$ is defined as an average:

$$E_t^f(p) = \frac{1}{\sum_{p' \in \mathcal{N}(p)} S(p, p')} \sum_{p' \in \mathcal{N}(p)} S(p, p').E_t(p'). \quad (25)$$

Finally, the binary decision is obtained by rounding off $D_t^f(p)$ when the final confidence level is great enough, and by setting its value to 1 (*foreground*) otherwise:

$$D_t^b(p) = \begin{cases} 0 \ (\textit{bgd}) & \text{if } D_t^f(p) \leq 0.5 \text{ and } E_t^f(p) > k_e \\ 1 \ (\textit{fgd}) & \text{if } D_t^f(p) > 0.5 \text{ and } E_t^f(p) > k_e \\ 1 & \text{if } E_t^f(p) \leq k_e. \end{cases} \quad (26)$$

After this final decision, a binary mask can be computed on the whole frame, to represent the locations of the *foreground* pixels. Such results are presented in the next section.

### *C. GMM update*

We choose to maintain the update processings proposed by Stauffner [7] (cf. II-C3). Only difference lies in the prior pixel mode selection: in our case, the decision does not require a mode selection (the decision is averaged over the whole pixel neighborhood).

So, we select the mode with *equal label* (foreground or background) and with maximal (crisp) likelihood - if greater than threshold $k_r$. If this selection fails, a new mode distribution replaces the lowest one, as above.

## IV. RESULTS

The proposed method is tested on two videos:

- *University video*: this is a real video recorded in the university, with a frame size of $480 \times 640$ pixels and a frame rate of 29 frames/s. The background is quite motionless, but the natural lighting changes.
- *Fountain video*: this video is taken from the *Change Detection 2014 benchmark dataset* [13] (1184 frames), with

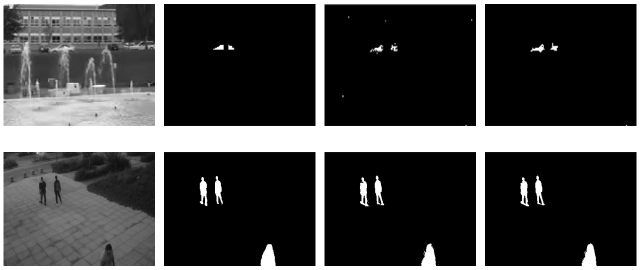

Figure 6. Original frames extracted of *Fountain* and *University* videos (Col. 1). Corresponding foreground ground truth (Col. 2). Foreground detected by the proposed method with the spatial fusion (Col. 3). Foreground detected by the proposed method without the spatial fusion (Col. 4).

a frame size of 160*128. The background is dynamic, due to the four fountains.

The algorithm is implemented in MATLAB® and the simulation is processed on an Intel® Core i7-4500U @ 1.80 GHz-2.40 GHz laptop with 16 GB RAM.

First, qualitative results obtained a single frame for each video sequence are shown on Figure 6. Moving objects are correctly identified.

Then, the F-measures are processed on both sequences: the obtained values are clearly promising (cf. Table I).

Table I
F-SCORES OF THE PROPOSED METHOD.

| Video | F-score |
|---|---|
| University video | 0.9322 |
| Fountain video | 0.8351 |

We can conclude that the final decisions are very relevant: moving objects are clearly identified (cars in the upper video, pedestrians in the lower one). There is almost no false detection because of both the uncertainty handling and of the spatial fusion step, which can perfectly manage such simple surveillance videos.

## V. Conclusion

In this work, we proposed a background - and foreground - modelling, based on an original type-2 fuzzy Gaussian Mixture Model. Intermediate fuzzy results are exploited through a spatial fusion step, using interval comparisons. Results on videos with dynamic backgrounds show that the method can adjust to such changing background, achieving a robust detection of the objects. This work shows the importance of fuzzy models to deal with uncertainty in the task of background subtraction. The next step is to compare this method with other (fuzzy or not) background subtraction methods.

## References


[1] A. R. François and G. G. Medioni, "Adaptive color background modeling for real-time segmentation of video streams," in *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, vol. 1, no. 121, 1999, pp. 227–232.

[2] W. I. Guerra and E. García-Reyes, "A novel approach to robust background subtraction," in *Iberoamerican Congress on Pattern Recognition*. Springer, 2009, pp. 69–76.

[3] M. Piccardi, "Background subtraction techniques: a review," in *Systems, man and cybernetics, 2004 IEEE international conference on*, vol. 4. IEEE, 2004, pp. 3099–3104.

[4] S. Y. Elhabian, K. M. El-Sayed, and S. H. Ahmed, "Moving object detection in spatial domain using background removal techniques-state-of-art," *Recent patents on computer science*, vol. 1, no. 1, pp. 32–54, 2008.

[5] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.

[6] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 780–785, 1997.

[7] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, vol. 2. IEEE, 1999, pp. 246–252.

[8] T. Bouwmans and F. El Baf, "Modeling of dynamic backgrounds by type-2 fuzzy gaussians mixture models," *MASAUM Journal of of Basic and Applied Sciences*, vol. 1, no. 2, pp. 265–276, 2009.

[9] A. Shimada, H. Nagahara, and R.-i. Taniguchi, "Background modeling based on bidirectional analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1979–1986.

[10] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection-a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, 2008.

[11] M. Chen, Q. Yang, Q. Li, G. Wang, and M.-H. Yang, "Spatiotemporal background subtraction using minimum spanning tree and optical flow," in *European Conference on Computer Vision*. Springer, 2014, pp. 521–534.

[12] J. Zeng, L. Xie, and Z.-Q. Liu, "Type-2 fuzzy gaussian mixture models," *Pattern Recognition*, vol. 41, no. 12, pp. 3636–3643, 2008.

[13] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "Cdnet 2014: An expanded change detection benchmark dataset," in *CDW-2014: Proceedings of IEEE Workshop on Change Detection at CVPR-2014*. New York, NY, USA: IEEE, 2014, pp. 393–400.

[14] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[15] R. Hosseini, J. Dehmeshki, S. Barman, M. Mazinani, and S. Qanadli, "A genetic type-2 fuzzy logic system for pattern recognition in computer aided detection systems," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–7.

[16] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *Fuzzy Systems, IEEE Transactions on*, vol. 7, no. 6, pp. 643–658, 1999.

[17] J. M. Mendel, "Type-2 fuzzy sets and systems: An overview [corrected reprint]," *IEEE Computational Intelligence Magazine*, vol. 2, no. 2, pp. 20–29, 2007.

[18] I. B. Turksen, "Interval valued fuzzy sets based on normal forms," *Fuzzy sets and systems*, vol. 20, no. 2, pp. 191–210, 1986.

[19] D. Dubois and H. Prade, "Interval-valued fuzzy sets, possibility theory and imprecise probability." in *EUSFLAT Conf.*, 2005, pp. 314–319.

[20] A. Bigand and O. Colot, "Membership function construction for interval-valued fuzzy sets with application to gaussian noise reduction," *Fuzzy Sets and Systems*, vol. 286, no. 1, pp. 66–85, 2016.

[21] J. Zeng and Z.-Q. Liu, "Type-2 fuzzy sets for handling uncertainty in pattern recognition," in *Fuzzy Systems, 2006 IEEE International Conference on*. IEEE, 2006, pp. 1247–1252.

[22] A. Sengupta and T. K. Pal, "On comparing interval numbers," *European Journal of Operational Research*, vol. 127, no. 1, pp. 28–43, 2000.