

Détection d'événements dans une séquence d'images avec caméra en mouvement

Rozenn DAHYOT¹, Pierre CHARBONNIER¹, Fabrice HEITZ²

¹Laboratoire Régional des Ponts-et-Chaussées de Strasbourg
11 Rue Jean Mentelin, BP 9, F-67035 STRASBOURG cedex

²LSIIT UPRESA CNRS 7005 Université Strasbourg I
Boulevard Sébastien Brandt F-67400 ILLKIRCH

Rozenn.Dahyot@equipement.gouv.fr, Pierre.Charbonnier@equipement.gouv.fr
heitz@lsiit.u-strasbg.fr

Résumé – La détection de changements dans les séquences d'images s'est principalement intéressée à la détection d'objets mobiles quand le système d'acquisition est statique, ou à la détection d'effets de production, comme les changements de plans. Lorsque la caméra est mobile, son mouvement est classiquement géré par compensation du mouvement dominant, ce qui met en œuvre des techniques d'estimation de mouvement et/ou de segmentation. Dans cet article, nous proposons une nouvelle méthode de détection de changements statistiques capable de gérer des événements complexes tels que l'entrée ou la sortie d'objets, et le changement d'apparence d'objets quand la caméra est en mouvement. Les changements temporels sont extraits en analysant les distributions statistiques d'images successives. Si l'on considère des mesures appropriées, nous montrons comment extraire les statistiques des objets changeants en utilisant deux histogrammes d'images successives. Ces objets sont ensuite localisés par une technique de rétroprojection. La méthode est complètement non supervisée et ne nécessite ni estimation, ni compensation du mouvement. Elle est illustrée sur des images de scènes routières présentant de grands mouvements de caméra.

Abstract – Change detection in image sequences has mainly focused on the recovery of moving objects when the viewing system is static, or on the detection of simple production effects such as video shot boundaries or scene transitions. Camera motion is usually handled by the compensation of dominant motion, using motion estimation and segmentation schemes. In this paper we propose a novel statistical change detection method able to handle more complex events such as entering or exiting objects, or changes in objects appearance, when the camera is moving. Temporal changes of objects are captured by analyzing the statistics of successive images. Considering an appropriate choice of image features, we show how it is possible to extract the statistics of changing objects from a pair of successive image histograms. Changing objects are then located by a back-projection technique. The method is completely unsupervised and does not require motion estimation nor motion compensation. It is illustrated here on real world road scenes exhibiting large camera motion.

1 Introduction

L'analyse automatique de séquences vidéos est une tâche essentielle dans des applications telles que l'indexation par le contenu et le codage. Dans cet article, nous proposons une nouvelle méthode de détection des changements statistiques sensible aux variations temporelles complexes dans les séquences d'images acquises avec une caméra en mouvement. Des approches statistiques non supervisées se sont déjà montrées efficaces, rapides et robustes aux mouvements de caméra pour la détection de ruptures simples de scènes [6, 1]. Nous étendons ce concept à la détection d'événements complexes générés par la sortie, l'entrée ou le changement d'apparence d'objets dans la séquence.

Nous supposons que chaque image de la séquence est composée d'un ensemble d'objets *changeants* \mathcal{B} (objets d'intérêt à détecter), et d'un arrière-plan $\bar{\mathcal{B}}$ statistiquement peu changeant. En pratique, nous sommes intéressés par les objets sortants ou entrants dans la vidéo ou les objets dont les mesures locales changent significativement au cours du temps. A partir de quelques hypothèses sur \mathcal{B} et $\bar{\mathcal{B}}$, nous montrons dans la section 2.2 comment ex-

traire les densités de probabilité des objets changeants. Des méthodes statistiques de rétroprojection sont ensuite utilisées pour localiser les objets d'intérêt dans les images [5] (section 3).

Cette méthode de détection est non supervisée et rapide puisqu'elle ne nécessite ni recalage d'images, ni estimation de mouvement [2]. Elle est appliquée à la détection d'un ensemble d'objets dans les séquences de scènes routières prises par un véhicule en mouvement (section 4).

2 Détection des changements statistiques

On considère une séquence d'images $I(t)$ d'une scène composée de deux classes : l'arrière-plan $\bar{\mathcal{B}}$ qui constitue la majeure partie de la scène, et les objets d'intérêt \mathcal{B} . Tandis que les variations statistiques de l'arrière-plan sont supposées lentes, les objets changent fortement en taille ou en apparence. Ils peuvent même entrer ou sortir de la scène. Contrairement aux méthodes supervisées, on ne réalise pas d'apprentissage sur les statistiques des objets d'intérêt.

Cependant, nous supposons que nous pouvons définir une mesure locale \mathbf{m} qui permet de discriminer l'objet de son arrière-plan. Au paragraphe suivant, nous posons ces hypothèses sous forme mathématique. Puis nous montrons comment extraire les statistiques propres aux objets changeants $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$ à partir des statistiques de deux images successives $\mathcal{P}(\mathbf{m}, t)$ et $\mathcal{P}(\mathbf{m}, t+1)$ à l'aide d'«extracteurs» :

$$\begin{cases} \Delta_+(\mathbf{m}, t) = \max(\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1), 0) \\ \Delta_-(\mathbf{m}, t) = \max(\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t-1), 0) \end{cases} \quad (1)$$

où γ est un paramètre de «sensibilité».

2.1 Hypothèses

H_0 : Composition des images $I(t)$. On suppose que chaque image $I(t)$ de la séquence est composée d'un ensemble d'objets changeants \mathcal{B} et d'un arrière-plan $\bar{\mathcal{B}}$. La densité de probabilité $\mathcal{P}(\mathbf{m}, t)$ des mesures locales \mathbf{m} dans l'image $I(t)$ peut s'écrire :

$$\mathcal{P}(\mathbf{m}, t) = (1 - \epsilon(t)) \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B})$$

H_1 : Proportion des objets de \mathcal{B} . L'hypothèse H_1 restreint le domaine de définition de la proportion ϵ des objets et est liée au choix du paramètre γ dans les extracteurs. Nous supposons que ϵ appartient à l'intervalle $[0, \frac{\gamma-1}{\gamma}]$.

H_2 : Variations statistiques du fond $\bar{\mathcal{B}}$. On suppose que les statistiques du fond varient faiblement au cours du temps, soit :

$$\forall \mathbf{m}, \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \simeq \mathcal{P}(\mathbf{m}, t+1|\bar{\mathcal{B}})$$

H_3 : Différence statistique entre \mathcal{B} et $\bar{\mathcal{B}}$. Cette hypothèse suppose que l'on peut choisir une mesure locale \mathbf{m} qui permet de distinguer les objets de \mathcal{B} du fond $\bar{\mathcal{B}}$:

$$\forall \mathbf{m}, \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \cdot \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \simeq 0$$

H_4 : Variations statistiques de \mathcal{B} . En pratique, nous sommes intéressés par les objets dont les mesures locales varient significativement au cours du temps ou ceux qui apparaissent, changent de proportion ou disparaissent de la scène. Ainsi deux cas peuvent être envisagés. Le premier H_4 considère que $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$ change fortement au cours du temps :

$$\forall \mathbf{m}, \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \cdot \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \simeq 0$$

Le second cas H_4' suppose au contraire que :

$$\forall \mathbf{m}, \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \simeq \mathcal{P}(\mathbf{m}, t+1|\mathcal{B})$$

mais que c'est la proportion $\epsilon(t)$ de l'objet qui varie *fortement*.

2.2 Extracteurs

Lorsque les hypothèses H_0, H_1, H_2, H_3 et H_4 sont vérifiées entre les instants t et $t+1$, et entre t et $t-1$ alors (cf. annexe A) :

$$\begin{cases} \Delta_+(\mathbf{m}, t) = \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \\ \Delta_-(\mathbf{m}, t) = \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \end{cases} \quad (2)$$

Les densités de probabilité devant s'intégrer à 1, on peut facilement calculer $\epsilon(t)$ et extraire la distribution des objets changeants $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$. On note que le choix de γ n'a aucune influence sur l'extraction quand H_4 est vérifiée. Dans le cas où c'est l'hypothèse H_4' qui est vérifiée, alors (cf. annexe A) :

$$\begin{cases} \Delta_+(\mathbf{m}, t) = \max(\epsilon(t) - \gamma \epsilon(t+1), 0) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \\ \Delta_-(\mathbf{m}, t) = \max(\epsilon(t) - \gamma \epsilon(t-1), 0) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \end{cases} \quad (3)$$

Le paramètre de sensibilité γ permet de contrôler l'amplitude des variations statistiques que l'on souhaite détecter. Dans le cas de l'hypothèse H_4' , Δ_+ (resp. Δ_-) détecte les objets dont la proportion diminue au moins d'un facteur γ (resp. augmente au moins d'un facteur γ).

3 Localisation des changements statistiques

Pour localiser les objets de \mathcal{B} dans $I(t)$, nous utilisons la technique de rétro-projection bayésienne [5]. A chaque position (x, y) de l'image $I(t)$, une mesure locale $\mathbf{m}_{x,y}$ est calculée et sa probabilité *a posteriori* $\mathcal{P}(\mathcal{B}|\mathbf{m}_{x,y}, t)$ d'appartenir à \mathcal{B} , calculée grâce à la formule de Bayes, est reportée dans une carte de confiance :

$$\mathcal{P}(\mathcal{B}|\mathbf{m}_{x,y}, t) = \frac{\mathcal{P}(\mathbf{m}_{x,y}, t|\mathcal{B}) \mathcal{P}(\mathcal{B})}{\mathcal{P}(\mathbf{m}_{x,y}, t)}$$

La probabilité *a priori* $\mathcal{P}(\mathcal{B})$ n'intervient pas dans la décision. Pour chaque image $I(t)$ de la séquence, deux cartes complémentaires de rétro-projection sont calculées, l'une dite *en arrière* calculée grâce à Δ_+ , et l'autre dite *en avant* calculée grâce à Δ_- .

La procédure de pré-détection est résumée à la figure 1 : les densités de probabilité de trois images successives sont calculées grâce à des histogrammes. Puis, les détecteurs Δ_+ et Δ_- sont calculés et les rétroprojections en avant et en arrière dans l'image $I(t)$ sont réalisées.

4 Résultats expérimentaux

Dans cette expérience, on considère une séquence de 1000 images couleurs de scènes routières filmées par une caméra montée sur un véhicule. Une image est prise tous les 5 mètres : le mouvement de la caméra correspond à un travelling avant relativement important. Nos objets d'intérêt sont les obstacles fixes présents sur les bas-côtés de la route comme les panneaux, les délinéateurs, les poteaux, les arbres, etc... Ces objets apparaissent, grossissent et disparaissent dans la séquence. Pour respecter l'hypothèse H_3 , deux mesures locales ont été choisies :

- la première utilise l'information couleur, communément utilisée dans les applications de détection de panneaux routiers,
- la seconde est une mesure de forme qui permet de caractériser les alignements, très présents dans nos objets d'intérêt.

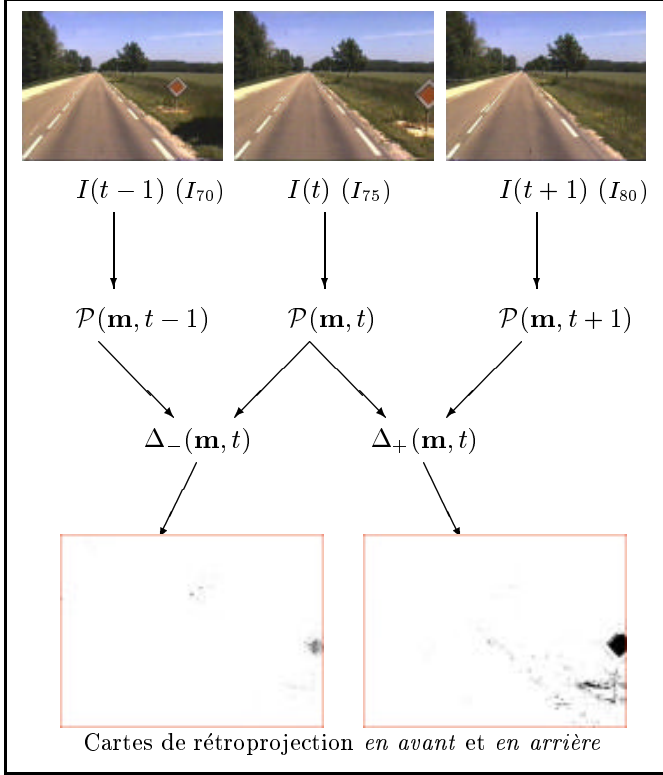


FIG. 1: **Détection des objets changeants** ($\gamma = 2$): Les mesures (couleur) “changeantes” apparaissent en sombre dans les cartes de rétroprojection de I_{75} (voir texte).

4.1 Mesure de couleur

Les mesures de couleur choisies sont les chromaticités (r, g) extraites des valeurs RGB des pixels :

$$\mathbf{m}_{couleur} = \begin{pmatrix} r = \frac{R}{R+G+B} \\ g = \frac{G}{R+G+B} \end{pmatrix} \quad (4)$$

La figure 1 présente les cartes de rétroprojection obtenues sur l'image I_{75} , avec les mesures de couleur. Les pixels blancs indiquent une probabilité nulle tandis que les pixels noirs correspondent à une grande probabilité. La proportion de pixels de probabilité nulle est très importante, ce qui indique que la méthode est très discriminante. C'est pourquoi la carte de rétroprojection n'est pas seuillée: toutes les positions de probabilité non nulle sont retenues ici. Le panneau jaune grossit entre les images I_{70} et I_{75} . Sa couleur ne changeant pas, l'hypothèse H_4 est vérifiée: la proportion ϵ augmente d'un facteur supérieur à γ , ce qui provoque la détection par la rétroprojection en avant dans I_{75} . Entre les images I_{75} et I_{80} , le panneau disparaît. H'_4 est encore vérifiée: la proportion s'annule ce qui permet de détecter le panneau en rétroprojection en arrière dans I_{75} .

4.2 Mesure de forme

Trois paramètres définissent la mesure locale de forme, caractérisant les alignements. Le premier correspond à l'angle θ du contour local. Le second, α , est une mesure d'alignement: deux mesures locales appartenant à un même contour rectiligne ont des valeurs (θ, α) identiques.

Le troisième paramètre est la norme du gradient :

$$\mathbf{m}_{forme} = \begin{pmatrix} \theta = \arg \tan \frac{I_x}{I_y} \\ \alpha = x \frac{I_x}{N} + y \frac{I_y}{N} \\ N = \sqrt{I_x^2 + I_y^2} \end{pmatrix} \quad (5)$$

où I_x et I_y correspondent aux dérivées premières de l'image d'intensité I . La figure 2 montre les cartes de rétroprojection obtenues par notre méthode. Les contours disparais-

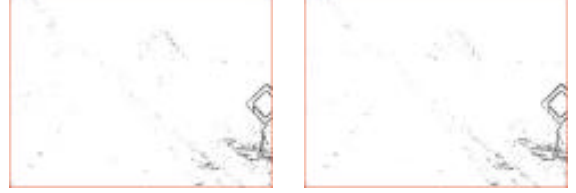


FIG. 2: Cartes de rétroprojection en avant et en arrière de l'image I_{75} , calculées avec des mesures de forme.

sants du panneau sont parfaitement détectés par la rétroprojection en arrière (hypothèse H'_4). La valeur de α est sensible à la position (x, y) dans l'image. Ainsi, la détection des contours du panneau dans la rétroprojection en avant est due aux changements des statistiques des mesures de forme du panneau (hypothèse H_4). Comme on peut le remarquer sur la figure 3, notre méthode est évidemment différente d'une détection de contours. En effet, seuls les contours relatifs à des objets statistiquement *changeants*, comme le délinéateur, sont détectés. Au contraire, les statistiques relatives aux marquages routiers restent les mêmes.

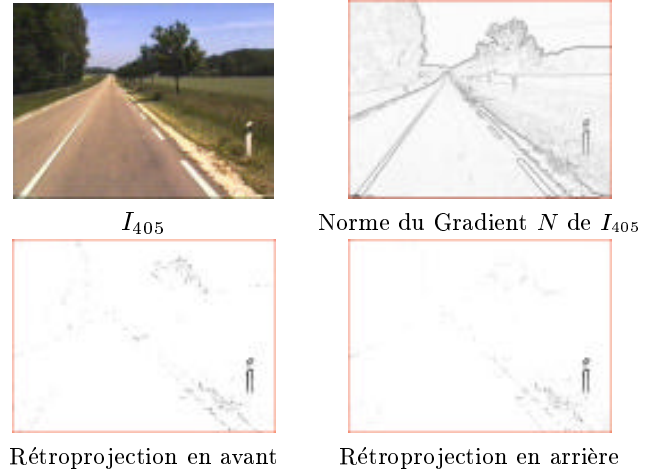


FIG. 3: Résultats de détection pour l'image I_{405} avec des mesures de forme.

4.3 Résultats expérimentaux

Fusion des cartes de rétroprojection en arrière et en avant. Les cartes de rétroprojection *en avant* et *en arrière*, complémentaires, sont sommées. Les régions de taille inférieure à 10 pixels, sont éliminées par une analyse en composantes connexes. La figure 4 présente les résultats de fusion pour l'image I_{75} . La valeur $\gamma = 2$ a été choisie

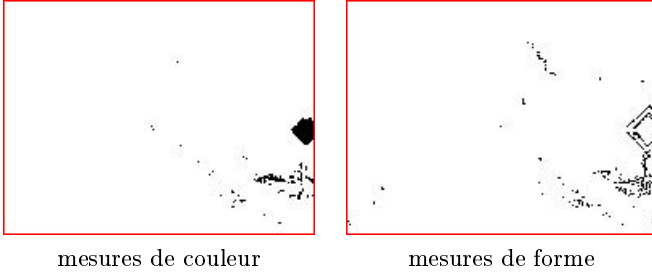


FIG. 4: Fusion des cartes *en avant* et *en arrière* pour l'image I_{75} .

pour cette expérience car elle est un bon compromis entre la sensibilité des détecteurs à des changements d'échelle (hypothèse H_4') et la proportion maximum attendue des objets dans les images (hypothèse $H_1 : \epsilon \in [0, \frac{1}{2}]$).

Résultats expérimentaux. Dans la séquence de 1000 images, il y a 21 panneaux routiers, 78 délinéateurs et 54 troncs d'arbre. On considère qu'un objet est détecté si ses contours ou régions colorées sont détectés dans au moins une des images de la séquence où il apparaît. Le tableau 1 présente les résultats de détection obtenus avec les mesures de forme et de couleur. Pour la recherche des troncs d'arbres et des délinéateurs, seules les statistiques verticales ont été utilisées. Les panneaux manqués par la détection avec les mesures de couleur, soit présentent peu de régions colorées, soit sont situées dans des zones fortement ombragées. Pour la détection avec les mesures de forme, les oublis sont dus principalement aux ombres qui diminuent fortement l'information de contour (gradient faible).

TAB. 1: Résultats de pré-détection.

Mesure	Couleur	Forme
Panneau (21)	57%	81%
Délinéateurs (78)	-	93%
Troncs d'arbres (54)	-	98%

5 Conclusion

Nous avons montré que l'information statistique relative aux objets changeants pouvait être extraite dans les séquences d'images prises par une caméra en mouvement. La méthode de détection est rapide, non supervisée, et peut être adaptée à la détection de différents objets en sélectionnant une mesure locale adaptée.

Références

- [1] J.S. Boreczky and L.A. Rowe. Comparison of video shot boundary detection techniques. *In Storage and Retrieval for Image and Video Databases IV, proc. SPIE 2670*, pages 170–179, 1996.
- [2] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: A synopsis of current problems

and methods. *Int. Jl. Computer Vision*, 19(1):29–55, 1996.

- [3] B. Schiele and A. Pentland. Probabilistic object recognition and localization. In *proc. of International Conference of Computer Vision ICCV'99*, 1999.
- [4] H. Schneiderman and T. Kanade. A statistical method for object detection applied to faces and cars. In *proc. of conference of Computer Vision and Pattern Recognition CVPR'00*, volume 1, Hilton Head, USA, June 2000.
- [5] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
- [6] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying production effects. *Multimedia systems*, 7:119–128, 1999.

A Calcul de $\Delta_+(\mathbf{m}, t)$

On dispose de deux distributions $\mathcal{P}(\mathbf{m}, t)$ et $\mathcal{P}(\mathbf{m}, t+1)$, des mesures \mathbf{m} calculées sur deux images $I(t)$ et $I(t+1)$. Grâce à H_0 , on a :

$$\begin{aligned} \mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1) = \\ (1 - \epsilon(t)) \mathcal{P}(\mathbf{m}, t|\overline{\mathcal{B}}) - \gamma (1 - \epsilon(t+1)) \mathcal{P}(\mathbf{m}, t+1|\overline{\mathcal{B}}) \\ + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) - \gamma \epsilon(t+1) \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \end{aligned} \quad (6)$$

Par H_2 , $\mathcal{P}(\mathbf{m}, t+1|\overline{\mathcal{B}}) \simeq \mathcal{P}(\mathbf{m}, t|\overline{\mathcal{B}})$ et l'équation (6) devient :

$$\begin{aligned} \mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1) = \\ ((1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))) \mathcal{P}(\mathbf{m}, t|\overline{\mathcal{B}}) \\ + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) - \gamma \epsilon(t+1) \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \end{aligned} \quad (7)$$

Cas de H_4 . H_3 et H_4 impliquent que si l'un des trois termes $\mathcal{P}(\mathbf{m}, t|\overline{\mathcal{B}})$, $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$ et $\mathcal{P}(\mathbf{m}, t+1|\mathcal{B})$, est non nul, alors les autres sont nuls. Donc :

$$\begin{aligned} |\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1)| = \\ |(1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))| \mathcal{P}(\mathbf{m}, t|\overline{\mathcal{B}}) \\ + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) + \gamma \epsilon(t+1) \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \end{aligned} \quad (8)$$

D'après H_1 , $(1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1)) < 0$ d'où :

$$|1 - \epsilon(t) - \gamma (1 - \epsilon(t+1))| = \epsilon(t) - 1 + \gamma (1 - \epsilon(t+1)) \quad (9)$$

Comme $\max(a, 0) = \frac{a+|a|}{2}$, on a :

$$\Delta_+(\mathbf{m}, t) = \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \quad (10)$$

Cas de H_4' . Dans ce cas $\mathcal{P}(\mathbf{m}, t|\mathcal{B}) = \mathcal{P}(\mathbf{m}, t+1|\mathcal{B})$, l'équation (7) devient :

$$\begin{aligned} \mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1) = \\ ((1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))) \mathcal{P}(\mathbf{m}, t|\overline{\mathcal{B}}) \\ + (\epsilon(t) - \gamma \epsilon(t+1)) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \end{aligned} \quad (11)$$

et avec H_3 :

$$\begin{aligned} |\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1)| = \\ |(1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))| \mathcal{P}(\mathbf{m}, t|\overline{\mathcal{B}}) \\ + |\epsilon(t) - \gamma \epsilon(t+1)| \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \end{aligned} \quad (12)$$

Donc,

$$\Delta_+(\mathbf{m}, t) = \max(\epsilon(t) - \gamma \epsilon(t+1), 0) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \quad (13)$$