

①②

DEMANDE DE BREVET D'INVENTION

A1

②② Date de dépôt : 03.10.00.

③⑦ Priorité :

④③ Date de mise à la disposition du public de la
demande : 05.04.02 Bulletin 02/14.

⑤⑥ Liste des documents cités dans le rapport de
recherche préliminaire : *Ce dernier n'a pas été
établi à la date de publication de la demande.*

⑥⑦ Références à d'autres documents nationaux
apparentés :

⑦① Demandeur(s) : *TELEDIFFUSION DE FRANCE TDF
Société anonyme* — FR.

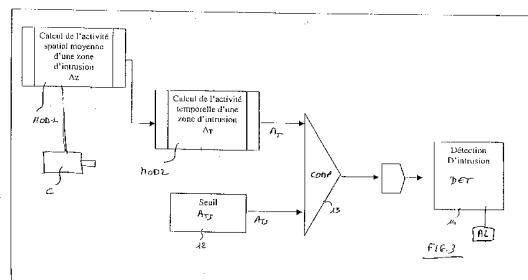
⑦② Inventeur(s) : BAINA JAMAL et GERARDIN OLI-
VIER.

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : CABINET ORES.

⑤④ PROCÉDE DE DETECTION D'INTRUSION DANS UN SYSTEME DE TELESURVEILLANCE.

⑤⑦ L'invention concerne un procédé de détection d'intrusion dans un système de télésurveillance mettant en oeuvre au moins une caméra vidéo fournissant une image vidéo, caractérisé en ce qu'il met en oeuvre pour une image résultante obtenue en soustrayant les valeurs pixel par pixel d'une image courante et d'une image de référence, une transformée par blocs, dans au moins une zone caractéristique de l'image comprenant N blocs, puis le calcul d'une activité spatiale moyenne des n blocs de ladite zone, notée Az et qui constitue un signal numérique de détection d'intrusion.



PROCEDE DE DETECTION D'INTRUSION DANS UN SYSTEME DE TELESURVEILLANCE

L'invention concerne un procédé de détection d'intrusions dans un système de télésurveillance mettant en œuvre au moins une caméra.

5 La télésurveillance vidéo est utilisée dans différents domaines d'activité notamment en ce qui concerne la surveillance des biens et des personnes : surveillance de sites industriels, surveillance de lieux publics, surveillance du trafic routier.

L'invention s'intéresse plus particulièrement à la surveillance
10 de lieux afin de détecter les intrusions à l'aide d'une caméra, notamment une caméra fixe. Dans l'état de la technique, il est connu d'utiliser des caméras fixes ou mobiles reliées à un pupitre d'écrans de contrôle qu'un opérateur regarde en permanence. Dans la pratique, les opérateurs n'ont pas leur attention fixée sur les écrans en permanence. De plus, un tel système
15 nécessite d'avoir un nombre d'écrans aussi important que le nombre de caméras.

Une solution connue consiste à afficher sur un nombre réduit d'écrans, de manière séquentielle, ce que filme chaque caméra. Cette solution permet de réduire le nombre d'écrans et de rendre la surveillance des
20 écrans moins monotone afin d'augmenter la vigilance de la personne qui surveille les écrans.

Une amélioration consiste à doubler chaque caméra d'un détecteur de présence ou d'événement afin d'afficher de manière automatique les images d'une caméra dont le détecteur détecte un mouvement sur l'image.
25 L'opérateur visualise alors à l'aide d'un écran de contrôle quelle est la situation. L'appareil peut également déclencher un enregistrement des images des caméras ayant détecté l'événement.

Afin de supprimer les détecteurs associés à chaque caméra, il est possible d'utiliser un traitement d'image pour détecter les mouvements sur
30 l'image par rapport à une image de référence. Classiquement, cela se fait pour des caméras fixes qui filment en permanence un même lieu selon un angle unique. Une image de référence correspondant au lieu à surveiller est numérisée et chaque image filmée est également numérisée puis comparée à l'image de référence.

35 Le domaine du codage d'image numérique a pris une importance considérable depuis les années 1980. Parmi les techniques

utilisées, certaines sont dédiées au codage à réduction de débit (normes H261, JPEG, MPEG1, MPEG2, JPEG2000, etc ...) et se basent toutes sur le principe de codage par transformation de blocs. Grâce à cette transformation, les algorithmes peuvent calculer des paramètres de corrélation spatiale, de corrélation temporelle, de redondance subjective ou de redondance statistique afin de comprimer fortement les images. Parmi les transformées connues, la transformée en cosinus discrète (par la suite TCD) et la transformée en ondelettes sont les plus utilisées afin de déterminer la corrélation spatiale d'une image.

Ces algorithmes utilisent également des techniques de prédiction pour réduire davantage la taille des images compressées. Il s'agit pour ces algorithmes (comme les normes MJPEG, MPEG1, MPEG2, JPEG2000) de pouvoir prédire pour une image à un instant donné ce qui aura changé dans l'image à l'instant suivant afin d'effectuer par exemple de la compensation de mouvement. Ces techniques de prédiction se basent sur l'extraction des différences entre deux images en effectuant une soustraction desdites images et une estimation de mouvement. Afin de maintenir une qualité d'image acceptable, le prédicteur est en permanence remis à jour (par des techniques de compression puis décompression). Dans ce cas, on parle d'image « I codée en intra ou de référence instantanée » (prédicteur de base), d'image « B non prédictive » (n'intervenant pas dans la prédiction) et d'image « P prédictive » (prédiction temporelle).

Les systèmes de détection se basent classiquement sur la prise d'une image de référence de la scène filmée puis sur la comparaison de l'image filmée avec ladite image de référence par différentiation point par point (c'est-à-dire pixel par pixel) de l'image filmée et de l'image de référence pour obtenir une image de différence et enfin déclenchement d'une alarme si l'image contient des éléments mobiles, c'est-à-dire des différences constatées sur plusieurs images successives, et affichage de l'image sur un écran de contrôle pour que l'opérateur contrôle visuellement le lieu surveillé.

Un tel système présente quelques défauts :

- un premier défaut vient du fait que le lieu surveillé doit être éclairé d'une manière homogène, ce qui convient très bien pour les lieux éclairés à l'aide de lumière artificielle (par exemple une salle sans fenêtre) ou de manière indirecte (par exemple un couloir éclairé par réflexion de lumière sur un mur). Par contre, si le lieu à surveiller est éclairé de manière directe par

la lumière du soleil ou s'il s'agit d'un extérieur, de nombreuses fausses alarmes peuvent se déclencher, réduisant ainsi la vigilance de l'opérateur qui ne va plus pouvoir contrôler les écrans dont les alarmes se déclenchent en permanence ;

5 - un deuxième défaut vient également des fausses alarmes qui peuvent être déclenchées par un changement de décor d'un bureau (affiche, plante ...) ;

 - un troisième inconvénient provient des fausses alarmes dues à un seuil de référence trop bas par rapport à la qualité de l'image ;

10 - un quatrième défaut est que ces techniques ne s'intègrent pas facilement aux procédés de codage d'image à réduction de débit.

L'invention propose une solution qui permet notamment, de réduire de manière notable les fausses alarmes en extrayant de l'image (ou d'une partie de l'image) une signature permettant la détection.

15 Cette signature constitue une représentation temporelle de l'image, relativement peu sensible, aux changements de luminosité ou aux changements de décors mais fortement sensible aux intrusions, notamment lorsqu'elles se manifestent par une variation de contenu de l'image, qui se poursuit au cours du temps.

20 D'autre part, l'invention propose d'élaborer cette signature à partir de techniques de transformation par blocs, notamment la transformation en cosinus discrète TCD. Il s'agit donc d'apporter une forte valeur ajoutée à ses techniques sans pour autant augmenter de manière importante le volume de calcul à effectuer.

25 Ces caractéristiques avantageuses font de ce système un procédé robuste, performant et efficace.

Selon un premier aspect, l'invention concerne un procédé de détection d'intrusion dans un système de télésurveillance mettant en œuvre au moins une caméra vidéo fournissant une image vidéo qui met en œuvre
30 pour une image résultante obtenue en soustrayant les valeurs, pixel par pixel d'une image courante et d'une image de référence, une transformée par blocs, dans au moins une zone caractéristique de l'image résultante comprenant N blocs, pour obtenir pour chaque bloc $B_{n,m}$ des coefficients transformés $F_{n,m}(i, j)$, i et j variant de 0 à M, M désignant le nombre maximal
35 de lignes ou de colonnes de pixels d'un bloc $B_{n,m}$, puis le calcul d'une activité spatiale moyenne A_z des N blocs de ladite zone, de l'image résultante,

laquelle constitue un signal numérique de détection d'intrusion qui est comparé avec un seuil prédéterminé de détection d'intrusion pour détecter une intrusion.

- Le procédé peut mettre en œuvre, pour le calcul de l'activité spatiale A_Z :
- le calcul de l'activité spatiale A_{su} de chacun des N blocs $B_{n,m}$, u variant de 0 à $N-1$
 - le calcul de l'activité spatiale moyenne A_Z avec :

$$A_Z = 1/N \sum_{u=0}^{N-1} A_{su}$$

L'activité spatiale A_{su} peut être par exemple calculée à l'aide de la formule (7) donnée plus loin dans la description.

- Le procédé peut mettre en œuvre une compensation de la variation de la luminosité. On a alors :

$$A_Z = \frac{1}{N} \sum_{u=0}^{N-1} C_{su} \text{ avec } C_{su} = \frac{p \sqrt{A_{su}}}{d_{su}} \text{ avec } p \text{ constante, par exemple avec } p \text{ entier } \geq 2$$

d_{su} représentant la luminosité moyenne du bloc de rang u .

- Avantageusement, le procédé met en œuvre l'application d'une transformée temporelle unidimensionnelle à N' signaux de détection $A_{Z0}, \dots, A_{ZM-1}$, calculés pour des images courantes prises à des instants $t_0, \dots, t_{N'-1}$, pour obtenir une matrice de coefficients transformés $D_0, D_1, \dots, D_{N'-1}$, et un calcul d'un signal A_T représentatif d'une évolution temporelle de l'image, avec :

$$A_T = \left(\sum_{u=1}^{N'-1} D_u^2 \right)$$

$$D_0 \text{ est proportionnel à } \frac{1}{N'} \sum_{u=0}^{N'-1} A_{Zu}$$

- Le procédé peut être caractérisé en ce qu'il met en œuvre la comparaison du signal d'évolution temporelle A_T avec un seuil et

déclenchement d'une action de détection d'intrusion si le signal A_T franchit ledit seuil.

Le procédé peut mettre en œuvre plusieurs signaux de détection A_Z correspondant à plusieurs zones caractéristiques distinctes de l'image. Au moins un signal de détection A_Z peut être élaboré à partir d'au
5 moins une zone caractéristique de l'image.

Sous son premier aspect, l'invention a ainsi pour objet un procédé de traitement d'une image vidéo courante dans lequel on produit un premier signal associé à tout ou partie de l'image (appelée zone
10 caractéristique de l'image) et représentatifs de caractères de fixité et de variation desdites zones à un instant donné par rapport à une image de référence instantanée. Pour au moins une zone, on peut alors produire au moins un deuxième signal représentatif de l'évolution du premier signal dans le temps.

On peut utiliser plusieurs sous-signaux successifs pour
15 élaborer le signal de variabilité temporelle de la zone de l'image à partir duquel la prise de décision sera effectuée. Ainsi l'invention propose de traiter l'image par blocs auxquels est associé un premier signal instantané : l'activité spatiale A_s représentative de la situation instantanée d'un bloc de l'image résultante, puis un traitement de l'image par zones auxquels est associé un
20 deuxième signal instantané : le signal de détection A_Z représentatif de la variation ou de la fixité instantanée d'une zone de l'image par rapport à une image de référence. L'invention permet ensuite de traiter les zones dans le temps en élaborant le signal final de variabilité temporelle des zones de
25 l'image.

Sous son deuxième aspect, l'invention concerne un procédé de détection d'intrusions dans un système de télésurveillance, convient particulièrement à la détection d'une intrusion se manifestant sous la forme d'un dépointage et/ou d'une obstruction d'une caméra.

La vidéosurveillance ne peut être efficace si le champ
30 d'observation de la caméra, ou de tout autre capteur de prise de vue utilisé, est modifié par une action sur le capteur de manière à rendre invisibles les événements se produisant dans la scène. Afin de détecter ce type d'événement, le procédé selon le deuxième aspect de l'invention se base sur
35 les modifications induites sur le contenu des images. Le procédé fonctionne

ainsi sur une scène quelconque, et ne nécessite pas la présence d'un objet ou d'un marqueur spécifique dans la scène.

Pour ce faire, le procédé opère de préférence en deux étapes. La première (optionnelle) est une phase d'apprentissage. Celle-ci vise à extraire de la scène des éléments caractéristiques, de manière à obtenir une connaissance pertinente, représentative d'une situation normale. Dans une seconde phase dite opérationnelle, le procédé exploite sa connaissance de la scène pour comparer la scène courante aux caractéristiques d'une scène normale, et ainsi effectuer une détection.

Le procédé de l'invention sous son deuxième aspect est utilisable à chaque fois qu'une scène quelconque est observée par une caméra pointant en permanence le même champ. Le procédé en lui-même est indépendant du système de codage de l'image utilisé, mais il s'intègre très facilement dans un système de codage à réduction de débit standard utilisant la Transformée en Cosinus Discrète TCD.

Dans un mode de réalisation préféré, la transformée en cosinus discrète TCD, recommandée dans les standards JPEG et MPEG, est exploitée. L'invention tire profit des propriétés avantageuses de la TCD pour mettre en évidence les variations spatiales et temporelles dans l'image liées à son contenu. Cette approche originale permet ainsi de se baser exclusivement sur le traitement du signal vidéo. L'invention peut mettre en œuvre toute autre transformée par blocs linéaire et orthogonale, par exemple, la transformée de Fourier Discrète, la transformée en ondelettes (DWT) dont les coefficients ont été réorganisés pour obtenir les blocs transformés, ou bien la transformée de Hadamard, ou de Slant, ou bien encore, la transformée en sinus discrète.

Selon son deuxième aspect, l'invention concerne un procédé mettant en œuvre au moins une caméra vidéo fournissant une image vidéo, ladite intrusion à détecter se manifestant notamment par un dépointage et/ou une obstruction de ladite caméra, caractérisé en ce qu'il met en œuvre :

- a) une transformation, à l'aide d'une transformée par blocs, des blocs d'au moins un ensemble de blocs témoins de l'image comprenant N blocs $B_{n,m}$ de dimensions $M \times M$ pixels pour obtenir pour chaque bloc $B_{n,m}$ des coefficients transformés $F_{n,m}(i, j)$, i et j variant de 0 à $M-1$;
- b) la détermination, pour chacun des N blocs $B_{n,m}$, à partir des coefficients transformés $F_{n,m}(i, j)$, d'une activité spatiale $bsa_{n,m}$,

- le calcul pour au moins une image courante $I(t)$ et une image antérieure $I(t')$ prises à des instants t et t' d'un facteur d'évolution de ladite activité spatiale $[bsa_{n,m}(t), bsa_{n,m}(t')]$, de chacun des N blocs $B_{n,m}$, pondéré d'un facteur représentatif de l'évolution de la luminosité d'au moins un bloc dudit ensemble de blocs, pour constituer pour chaque bloc dudit ensemble de blocs, un paramètre caractéristique d'une intrusion qui est comparé avec un seuil prédéterminé de détection d'intrusion pour détecter une intrusion.

Une détection d'intrusion peut être déclarée lorsqu'un nombre prédéterminé N' de blocs de l'ensemble de N blocs témoins atteint ou dépasse au moins un dit seuil prédéterminé.

Selon un premier mode de réalisation, le procédé est caractérisé en ce que b met en œuvre :

- ladite détermination pour chacun des N blocs $B_{n,m}$ de l'image courante $I(t)$ et de l'image antérieure $I(t')$ de ladite activité spatiale, ainsi que d'une activité spatiale moyenne $[SA(t), SA(t')]$, des blocs dudit ensemble de blocs ;

- le calcul pour l'image courante $I(t)$ et pour l'image antérieure $I(t')$ d'un facteur $k_{n,m}(t)$ d'évolution de ladite activité spatiale $bsa(t)$, de chacun desdits N blocs $B_{n,m}$,

- le calcul pour lesdites images successives d'un facteur $K(t)$ d'évolution de l'activité spatiale moyenne $\{SA(t), SA(t')\}$ de l'image courante $I(t)$ par rapport à l'image antérieure $I(t')$ dans ledit ensemble de blocs.

- la détermination pour chaque bloc $B_{n,m}$ dudit ensemble de blocs d'un paramètre $ratio_{n,m}(t)$ caractéristique d'une intrusion en pondérant ledit facteur d'évolution $k_{n,m}(t)$ de l'activité spatiale des blocs par ledit facteur d'évolution moyenne $K(t)$ de l'activité spatiale, lequel est représentatif de l'évolution de la luminosité.

En particulier, le procédé est caractérisé en ce que :

$$k_{nm}(t) = bsa_{n,m}(t)/bsa_{n,m}(t') \text{ et}$$

$$K(t) = SA(t)/SA(t')$$

$$ratio_{n,m}(t) = k_{n,m}(t)/K(t)$$

Ladite comparaison avec un seuil prédéterminé peut mettre en œuvre une évaluation du paramètre $ratio_{nm}(t)$ pour l'image courante $I(t)$ et du paramètre $ratio_{nm}(t')$ pour ladite image antérieure $I(t')$ et une comparaison

de l'évolution de ce paramètre entre ces deux images avec au moins un seuil donné de détection d'intrusion constituant un dit seuil prédéterminé de détection d'intrusion.

Selon un deuxième mode de réalisation, le procédé est caractérisé en ce que b met en œuvre :

- pour chaque bloc $B_{n,m}$ de ladite zone, le calcul d'un paramètre d'activité spatiale $bsa_{n,m}$;
- le calcul pour chaque bloc $B_{n,m}$ dudit ensemble de blocs d'un paramètre d'activité spatiale pondéré $C_{n,m}$, représentatif du contraste dudit bloc, et le calcul pour chacun desdits blocs $B_{n,m}$ de l'image courante $I(t)$ et de l'image antérieure $I(t')$ d'un facteur $k'_{n,m}(t)$ d'évolution de ladite activité spatiale pondérée, pour obtenir ledit facteur d'évolution de l'activité spatiale pour chacun desdits n blocs.

Le procédé peut être caractérisé en ce que :

15

$$c_{n,m}(t) = \frac{\sqrt[p]{bsa_{n,m}(t)}}{dc_{n,m}(t)}$$

avec $dc_{n,m}(t) = F_{n,m}(0,0)$ coefficient transformé de rang (0,0) dudit bloc $B_{n,m}$ de l'image $I(t)$, représentant la luminosité moyenne du bloc $B_{n,m}$, et p étant une constante (par exemple p entier ≥ 2).

20

et en ce que :

$$k'_{n,m}(t) = c_{n,m}(t)/c_{n,m}(t')$$

Le procédé peut mettre en œuvre une évaluation du paramètre $k'_{n,m}(t)$ pour l'image courante $I(t)$ et du paramètre $k'_{n,m}(t')$ pour l'image antérieure $I(t')$ et une comparaison de l'évolution de ce paramètre entre ces deux images $[I(t), I(t')]$ avec au moins un deuxième seuil donné qui constitue un dit seuil prédéterminé de détection d'intrusion.

Le procédé peut être caractérisé en ce que :

30

$$bsa_{n,m} = \left(\sum_{i=0}^7 \sum_{j=0}^7 (a(i,j) \cdot F_{n,m}(i,j))^p \right)^{1/p} - (a(0,0) \cdot F_{n,m}(0,0))^p$$

Avec $a(i, j)$ constantes de pondération et p étant une constante (par exemple p entier ≥ 2)

5 Selon un mode de réalisation préféré de la phase de détection d'intrusion, l'image antérieure $I(t')$ est l'image $I(t-1)$ qui précède immédiatement l'image courante $I(t)$, cette image antérieure étant remplacée par ladite image courante $I(t)$ lorsqu'une nouvelle image courante $I(t+1)$ apparaît et qu'un dit paramètre caractéristique d'intrusion n'a pas atteint ou franchi un dit seuil prédéterminé de détection d'intrusion.

10 Ceci permet rafraîchissement de l'image antérieure qui sert de base à la comparaison, tant que la situation reste normale. Par contre, dès qu'un seuil de détection est franchi, l'image antérieure est figée, ce qui permet d'améliorer le diagnostic.

15 Selon une mise en œuvre préférée, le procédé met également en œuvre une phase d'apprentissage pour sélectionner des blocs de l'image constituant ledit ensemble de blocs témoins.

20 Le procédé peut alors être caractérisé en ce que la phase d'apprentissage met en œuvre la sélection parmi l'ensemble des blocs $B_{n,m}$ de l'image d'un dit ensemble ou dictionnaire de blocs témoins, pour lesquels la valeur du facteur d'évolution de l'activité spatiale reste sensiblement constante dans un intervalle de temps donné au cours duquel intervient au moins un événement choisi parmi une variation brusque de luminosité, l'extinction ou l'allumage d'au moins un éclairage, et/ou le déplacement d'une personne dans le champ de la caméra, tant que la caméra n'est pas soumise

25 à une obstruction et/ou à un dépointage.
La phase d'apprentissage peut mettre en œuvre une sélection des blocs $B_{n,m}$ de l'image présentant des éléments orientés sensiblement horizontalement et/ou verticalement.

30 Le procédé peut être caractérisé en ce qu'il met en œuvre une détection d'obstruction lorsqu'une baisse de l'activité spatiale de l'image $SA(t)$ par rapport à l'activité spatiale moyenne moyennée sur une séquence d'images précédentes, dépasse un seuil donné de baisse d'activité et lorsque l'activité spatiale moyenne $SA(t)$ des blocs $B_{n,m}$ de l'ensemble des blocs témoins de l'image courante présente une valeur en dessous d'un seuil
35 d'activité donné.

D'autres caractéristiques et avantages de l'invention apparaîtront à la lecture de la description ci-après, donné à titre d'exemple non limitatif, en liaison avec les dessins dans lesquels :

- les figures 1 à 3 illustrent l'invention sous son premier aspect, les figures 1 et 2 représentant les modules de calcul des paramètres d'activité A_z et A_T définis ci-après, et la figure 3 donnant un exemple d'architecture de l'ensemble ;
- les figures 4 à 15 illustrent l'invention sous son deuxième aspect, à savoir :
 - les figures 4 et 5 sont des algorithmes ;
 - les figures 6 à 8 illustrent l'influence des critères SA, DC et CG définis ci-après ;
 - la figure 9 est un exemple d'algorithme d'apprentissage ;
 - La figure 10 est un algorithme qui détaille la sélection des blocs orientés ;
 - la figure 11 illustre sous forme d'algorithme, la logique de déclenchement d'alarme ;
 - la figure 12 est un exemple d'architecture
 - et les figures 13 à 15 illustrent respectivement un exemple d'évolution du nombre de blocs du dictionnaire au cours d'un apprentissage, une détection de dépointage de la caméra, et d'une détection d'obstruction de la caméra.

Les figures 1 et 2 détaillent respectivement pour l'invention selon son premier aspect l'élaboration première d'un signal A_z représentatif de la variation ou de la fixité instantanée d'une zone 6 de l'image, puis l'élaboration du signal final A_T par des modules MOD 1 pour le signal A_z et MOD2 pour le signal A_T . La figure 3 illustre la prise de décision selon le procédé sous son premier aspect, à partir de la représentation finale.

On peut distinguer sur la figure 1 une image vidéo courante $I(t)$ fournie par une caméra-vidéo-surveillance C et une image de référence I_{ref} . Ces images sont issues des techniques de codage numérique préalablement citées. L'image vidéo courante $I(t)$ stockée dans un registre 1 et l'image de référence I_{ref} stockée dans un registre 2 sont de taille identique, par exemple 256 x 256 points (ou pixels). Dans l'exemple préféré, les images vidéo courante $I(t)$ et I_{ref} sont divisées en 1024 blocs de $M \times M = 8 \times 8$ points.

Chaque point correspond à une donnée numérique qui correspond par exemple à un niveau de gris codé sur un nombre entier, par exemple sur 8 bits. L'image de référence I_{ref} est soustraite point par point à l'image vidéo courante $I(t)$ afin d'obtenir une image différence ou image
 5 résultante I_{res} qui est stockée dans un registre 3.

L'image résultante I_{res} qui représente donc la différence entre l'image instantanée et l'image de référence; subit ensuite une transformée à deux dimensions effectuée bloc par bloc afin d'obtenir une image transformée IT représentative des différences entre l'image vidéo courante $I(t)$ et l'image
 10 de référence I_{ref} . L'image transformée IT est représentée par une matrice MAT de coefficients C_{ij} qui est stockée dans un registre 4.

L'image transformée IT est réalisée bloc par bloc pour deux raisons, qui sont d'une part la limitation de la puissance de calcul et d'autre part en vue d'un traitement par bloc de l'image qui sera abordé ultérieurement.

15 Il est toutefois possible de réaliser une transformée globale de toute l'image qui sera assimilée à un unique bloc.

La transformée réalisée dans l'exemple préféré est une transformée en cosinus discrète (par la suite TCD). D'autres transformées linéaires et orthogonales sont utilisables, entre autres la transformée de
 20 Fourier spatiale discrète, la transformée de Hadamard, la transformée en ondelettes, etc... Cependant la TCD présente comme avantage d'obtenir un nombre de coefficients identique au nombre de points et permet l'extraction directe de la luminosité moyenne du bloc traité.

La même transformée pourra être réutilisée selon l'invention
 25 sur une pile temporelle de N' valeurs d'entrée.

Chaque bloc de l'image résultante est représenté par une matrice de coefficients C_{ij} obtenus en appliquant la formule connue suivante (transformée en cosinus discrète) :

$$C_{ij} = \frac{4f(i)f(j)}{M^2} \sum_{n_1}^{M-1} \sum_{n_2}^{M-1} x(n_1, n_2) \cos\left(\frac{2\pi(2n_1 + 1)i}{4M}\right) \cos\left(\frac{2\pi(n_2 + 1)j}{4M}\right)$$

- dans laquelle M correspond au nombre maximum de lignes et de colonnes de points d'un bloc ;

$$\text{avec } f(x) = \frac{1}{\sqrt{2}} \text{ si } x = 0 \text{ et } f(x) = 1 \text{ si } x \neq 0$$

- dans laquelle i et j sont les indices des coefficients du bloc et sont compris entre zéro et M - 1 ; et

- x(n1, n2) correspond à la valeur numérisée du point de coordonnée n1, n2 dans le bloc avant transformation.

Pour chaque bloc, on extrait un signal As représentatif de la variation ou de la fixité qui correspond à une activité spatiale As dudit bloc. Ce signal est stocké dans un registre 5.

L'activité spatiale As est une combinaison d'au moins certains coefficients Cij de la matrice MAT constituant l'image transformée IT et elle peut être par exemple définie comme la somme des carrés de tous les coefficients Cij de la matrice MAT issue de la transformée, notamment la TCD, hormis le coefficient C00 qui est représentatif de la luminosité moyenne du bloc de l'image résultante correspondant. Ceci permet de s'affranchir dans une large mesure des variations de la luminosité moyenne de l'image.

On obtient donc, pour l'exemple préféré, une imagerie de 32 x 32 valeurs représentant l'activité spatiale Asu de tous les blocs de l'image. Puisque selon cet exemple, on effectue une somme de carrés, on prendra garde à coder As sur un nombre de bits suffisants (12 bits pour l'exemple proposé).

Pour chacun des N blocs de la zone 7, qui est une zone caractéristique ou un ensemble de blocs de l'image dans laquelle on effectue la détection d'intrusion, on calcule la moyenne de l'activité spatiale de tous les blocs de cette zone 7. Chaque zone est représentée pour l'image courante par la valeur Az obtenue par la formule suivante :

$$Az = \frac{1}{N} \sum_{i=0}^{N-1} A_{su}$$

30

Puisque l'activité spatiale As est codée sur 12 bits, il convient de coder l'activité spatiale moyenne d'une zone sur 16 bits ou plus. On obtient

alors un signal représentatif de la variation ou de la fixité de la zone pour l'image courante obtenue par différence entre une zone de l'image courante et la même zone de l'image de référence.

Si ce signal est nul, c'est que tous les N blocs de la zone
5 considérée de l'image courante sont identiques à ceux, pour la même zone, de l'image de référence. Si ce signal est faible, c'est que les différences pour ladite zone entre l'image de référence et l'image courante sont minimales.

Généralement, il s'agit ici de la traduction dans le signal transformé du bruit de l'image ou de légers changements de luminosité. Par
10 contre, lorsque le signal est important, c'est que les différences entre l'image de référence et l'image courante pour ladite zone sont elles mêmes importantes, et sont susceptibles de correspondre à la détection d'une intrusion. Pour obtenir une détection d'intrusion améliorée, on empile pour chaque zone, les N' (par exemple N' = 8) dernières valeurs de l'activité
15 spatiale moyenne de ladite zone 9, obtenues pour les N' dernières images captées. Ces valeurs sont stockées dans un registre 9.

A partir de ces valeurs, on applique une transformée, notamment TCD à une dimension. Chaque zone d'intrusion est alors représentée par une matrice 10 de coefficients D_i obtenus en appliquant la
20 formule suivante :

$$D_u = c(t) \sqrt{\frac{2}{N'}} \sum_{t=0}^{N'-1} f(t) \cos \left[\frac{u(2t+1)}{2N'} \right]$$

$$\text{avec } c(t) = \frac{1}{\sqrt{2}} \text{ si } t = 0 \text{ et } c(t) = 1 \text{ si } t \neq 0$$

- 25 - dans laquelle N' correspond au nombre maximum de valeurs empilées ;
- dans laquelle u est l'indice des valeurs empilées (compris entre 0 et N'-1) ;
- dans laquelle f(t) est égal à la valeur de l'activité spatiale
30 moyenne A_z de la zone choisie de l'image de rang t.

On extrait alors de la matrice stockée dans un registre 10 un signal A_T représentatif du niveau de variation de ladite zone dans le temps, qui correspond à une activité temporelle A_T définie comme une combinaison, par exemple la somme des carrés, de tous les coefficients D_{it} de la matrice 10 issue de la transformée TCD unidimensionnelle, hormis le coefficient D_0 qui est représentatif de la valeur moyenne des $N' = 8$ dernières activités spatiales moyennes. On a par exemple :

$$A_T = \sum_{t=1}^{N'-1} D_{it}^2$$

10

On obtient donc pour chaque zone un signal A_T qui est stockée dans un registre 11 et qui représente les variations dans le temps de l'activité spatiale moyenne des blocs de la zone. Ce signal A_T est représentatif des différences entre les blocs des huit dernières images courantes et ceux de l'image de référence pour ladite zone.

15

Si ce signal est faible, c'est que la plupart des blocs de la zone ne sont que très peu différents de ceux de l'image de référence et que ces différences restent constantes dans le temps.

Si ce signal A_T est important, c'est qu'un ou plusieurs blocs de la zone sont passés d'un contenu proche de celui de l'image de référence à un contenu qui diffère beaucoup par rapport à celui de l'image de référence. La valeur numérique de ce signal sera d'autant plus important. Ce signal permet donc de quantifier les mouvements ayant lieu dans la zone dans le temps.

25

La figure 3 illustre la mise en œuvre de la détection d'intrusion. Le signal A_T est généré par les modules MOD1 et MOD2 à partir des images vidéo fournies par la caméra C.

Ce signal A_T est donc comparé par un comparateur COMP 13 à un seuil A_{TS} dont la valeur est stockée dans un registre 12 afin de décider dans un circuit de décision DET s'il s'agit d'un mouvement suffisamment important pour être signalé par le biais d'une alarme AL.

30

L'invention selon son premier aspect permet donc de signaler tout mouvement anormal ayant lieu dans une zone prédéfinie. En positionnant judicieusement une ou plusieurs zones dans l'image, l'invention permet de ne surveiller que les zones de passage obligé et minimise ainsi le nombre de

35

calcul à effectuer. Elle permet également d'écarter du champ de la détection des parties de l'image comportant un éclairage (lampe ou autre) dont l'extinction, ou l'allumage, peut être dû à d'autres causes qu'une intrusion. Précisons qu'il est également possible de superposer plusieurs zones en
5 partie ou totalement afin de maximiser la détection.

L'invention permet également de signaler une alarme lorsque la caméra est bougée (l'homme du métier parle de dépointage de la caméra) ou que l'objectif de la caméra est obstrué ou bougé (l'homme du métier parle d'obstruction du champ de la caméra). Dans un tel cas, le contenu de l'image
10 varie beaucoup d'une image à l'autre, mais cette variation ne se prolonge pas dans le temps.

Le système peut être entièrement réalisé de manière logicielle ou entièrement réalisé à l'aide de composants électroniques dédiés.

Le système peut être implanté sur un dispositif matériel. Il
15 utilise plusieurs processeurs de signal (dits DSP de l'anglais Digital Signal Processor) pour réaliser les fonctions de soustraction d'image, les transformations TCD par blocs et pour calculer les activités spatiales et temporelles.

Les comparateurs, calcul de moyenne et prises de décisions
20 sont effectués par un module sur le processeur DSP. Il est important de signaler que le système doit traiter les informations en temps réel ou à défaut au rythme minimal des images acquises et traitées par seconde et par caméra. En effet, puisque l'on cherche à détecter un mouvement, il faut que l'écart temporel entre deux images successives traitées ne soit pas trop
25 important afin de retenir ce mouvement dans l'image au moment de l'acquisition. Un rythme d'une image par seconde et par caméra est à cet égard suffisant.

L'invention présente une méthode et un système qui effectue de la détection automatique de présence en utilisant notamment des procédés
30 et modules de codage exploités pour la transmission numérique à bas débit, comme on en rencontre dans le domaine de la télésurveillance.

Selon son deuxième aspect, l'invention concerne un procédé qui est plus particulièrement adaptable à la détection du dépointage et/ou d'une obstruction d'une caméra de surveillance.

Le dépointage d'une caméra peut être défini comme une action sur la caméra ayant pour effet de modifier la zone de la scène surveillée (rotation de la caméra sur son support, déplacement).

5 L'obstruction de champ d'une caméra peut être définie comme une action d'occuper tout ou partie de l'image par un obstacle ou objet placé devant l'objectif d'une caméra. On parle respectivement d'obstruction totale ou partielle.

Les données d'entrée du procédé de détection selon le deuxième aspect de l'invention, sont une séquence d'images.

10 La figure 4 donne le synoptique du détecteur de dépointage et/ou d'obstruction de caméra. Une alarme de dépointage, et/ou d'obstruction partielle ou bien une alarme d'obstruction totale peut être déclenchée lorsque le procédé estime que les conditions de détection sont réunies.

15 Le procédé s'appuie sur le principe de subdivision de l'image en blocs de $M \times M$ pixels, dont certains sont des blocs témoins. Les blocs témoins sont définis d'une part par un contenu remarquable, incluant certains éléments caractéristiques, et d'autre part par l'invariance de ce contenu.

20 Les blocs témoins sont avantageusement sélectionnés lors d'une phase d'apprentissage, qui lorsqu'elle est mise en œuvre, a lieu dans des conditions normales, c'est-à-dire en l'absence d'une situation d'alarme. La position des blocs témoins est mémorisée, et cet ensemble est appelé un dictionnaire.

25 Lorsque la phase d'apprentissage est terminée, la phase d'exploitation démarre. Lorsque d'un dépointage ou d'une obstruction, un certain nombre des blocs témoins change de contenu. Une alarme est déclenchée sur la base du nombre de blocs témoins ayant changé de contenu.

30 L'algorithme s'appuie sur le principe d'invariance des blocs témoins dans l'image en l'absence d'une situation anormale. Lorsque d'un dépointage ou d'une obstruction, un certain nombre de ces blocs caractéristiques perdent leur stabilité, et un critère de décisions permet de détecter un changement du contenu d'un bloc.

L'algorithme est donc organisé en deux phases distinctes (figure 5) : phase d'apprentissage et phase d'exploitation.

Les paragraphes suivants détaillent respectivement les critères de détection d'un changement de contenu d'un bloc de l'image, la phase d'apprentissage, puis la phase d'exploitation.

Le critère de changement de contenu dans l'image se base sur un ensemble de paramètres caractérisant le contenu de celle-ci. L'outil utilisé pour calculer ces paramètres est une transformée linéaire et orthogonale telle que la transformée en Cosinus Discrète (TCD).

Dans le cas bidirectionnel et pour une image subdivisée en blocs de $M \times M$ pixels, la TCD fait correspondre à un bloc $M \times M$ pixels un bloc de $M \times M$ coefficients. Ces derniers expriment la représentation fréquentielle de l'amplitude de la luminance des pixels. La transformée TCD s'exprime par la formulation mathématique suivante :

$$F(u, v) = \frac{4 \cdot c(u) \cdot c(v)}{M^2} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f(x, y) \cos \left(\left(\frac{2x+1}{2M} \right) \pi u \right) \cos \left(\left(\frac{2y+1}{2M} \right) \pi v \right) \quad (1)$$

$$F(0,0) = \frac{2}{M^2} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f(x, y) \quad (2)$$

$$c(u) = 1/\sqrt{2} \quad \text{si } u = 0$$

$$c(u) = 1 \quad \text{si } u \neq 0$$

Pour $u, v, x, y = 0, \dots, M-1$; par exemple $M=8$.

et (u, v) = index de la fréquence spatiale horizontale et verticale,

(x, y) = index de position horizontale et verticale bloc de pixel

$f(x, y)$ = luminance du pixel au point (x, y) dans l'image.

$F(u, v)$ = coefficient transformé au point (u, v) .

Les paramètres de caractérisation du contenu spatio-temporel utilisent les coefficients TCD.

Une information spatiale est fournie à partir du domaine transformé TCD par le calcul de l'activité spatiale de l'image. Ce paramètre informe précisément sur les détails dans l'image (I_t) indexée à l'instant t . L'image (I_t) contient $(H \times W)$ blocs de pixels. La taille du bloc $f_{n,m}(x,y)$ est M^2 pixels.

$$I_t = \left\{ f_{n,m}(x,y) / x,y = 0..M-1 / n = 0..H-1 / m = 0..W-1 \right\} \quad (3)$$

Dans un premier temps la transformée TCD est appliquée à l'image entière. Le premier facteur représentatif du contenu de l'image, l'activité spatiale, est obtenu à partir de l'image transformée $DCT(I_t)$.

Un bloc image très actif spatialement a une représentation
5 conséquente dans le domaine transformé. Ses coefficients ont une amplitude et une distribution spectrale importantes.

$$DCT(I_t) = \{F_{n,m}(u,v) = DCT(f_{n,m}(x,y)) \mid u,v,x,y = 0..M-1 / n = 0..B..H-1 / m = 0..B..W-1\}$$

10

Où ($H \times W$) est le nombre de blocs de l'image.

Deux paramètres sont introduits, (bsa) et ($SA(I_t)$). Ils représentent respectivement l'activité spatiale intra-bloc et intra-image. Le contenu analysé est celui de chaque bloc de pixel $f_{n,m}(x,y)$ positionné à n,m dans l'image (I_t).

15 L'indicateur ($bsa_{n,m}$) tient compte des composantes alternatives du bloc de coefficients transformés $F_{n,m}(u,v)$.

$$bsa_{n,m} = bsa(f_{n,m}(x,y)) = \left(\sum_{i=0}^7 \sum_{j=0}^7 (F_{n,m}(i,j))^2 \right) - (F_{n,m}(0,0))^2 \quad (5)$$

20 $SA(I_t)$ également noté $SA(t)$ est l'expression du contenu global de l'image $I(t)$ à l'instant t . Il est calculé par une moyenne des ($bsa_{n,m}$) concernés.

$$SA(I_t) = \overline{bsa_{n,m}(I_t)} = \frac{1}{N \times M} \sum_{n=0}^{H-1} \sum_{m=0}^{W-1} bsa_{n,m} \quad (6)$$

Où ($N \times M$) est le nombre des pixels de l'image (en considérant qu'au début de l'apprentissage, tous les blocs sont des blocs témoins).

25 L'indicateur $SA(I_t)$ permet de qualifier la richesse de l'image. Ce paramètre est affecté par le codage à réduction de débit et par les erreurs de transmission.

30 A titre d'illustration, on réalise la capture d'images prises avec éclairage artificiel (a), juste après extinction de la lumière (phase transitoire de réadaptation de la caméra (b), un peu plus tard par rapport à l'extinction de

l'éclairage (lorsque la caméra s'est réadaptée) (c), puis après allumage (d) et après réadaptation (e) suite à l'allumage. Si on calcule l'activité spatiale $bsa_{n,m}$ des blocs de ces images, on constate que l'activité spatiale varie en fonction de la luminosité.

5 La figure 6 montre l'évolution temporelle de l'activité spatiale bloc moyenne (SA) sur une séquence de 290 images, affectée par les évènements a à e précités, puis par un événement de dépointage/obstruction à partir de l'image 260. On repère nettement les phases d'extinction (b) et d'allumage (d) de l'éclairage artificiel, et de réadaptation (c) et (e). Par ailleurs,
10 on peut également constater une baisse ou une hausse transitoire de SA juste après la variation brutale de luminosité. Cela correspond au délai de réadaptation de la caméra C aux nouvelles conditions.

Les paramètres d'activité spatiale sont donc sensibles au contenu de l'image, mais également aux variations rapides et/ou importantes
15 de luminosité.

La définition de l'activité spatiale intra-bloc bsa choisie au paragraphe précédent peut être généralisée à toute formule utilisant les coefficients de la TCD. Cette considération est également valable pour l'activité spatiale As ou A_{su} définie ci-dessus dans le cadre de l'invention sous
20 son premier aspect.

$$hsa_{n,m} = \left(\sum_{i=0}^7 \sum_{j=0}^7 (a(i,j) \cdot F_{n,m}(i,j))^p \right) - (a(0,0) \cdot F_{n,m}(0,0))^p \quad (7)$$

Où $a(i,j)$ est un coefficient constant de pondération de la composante utilisée, et p une constante (par exemple p entier ≥ 2)

25 Plus largement, une activité spatiale peut être calculée à partir des résultats de toute transformation d'image linéaire et orthogonale appliquée bloc par bloc d'image (transformée en ondelettes, de Haar, de Hadamard, ...).

Une autre utilisation des coefficients transformés, permet de
30 définir un type de paramètre appelé contraste bloc local $c_{n,m}$, sensible au contenu de l'image mais peu aux variations de luminosité.

On utilise pour cela le coefficient (0,0) de la transformée, par exemple TCD, par exemple de chaque bloc :

$$dc_{n,m} = F_{n,m}(0,0) \quad (8)$$

$dc_{n,m}$ représente la luminosité moyenne du bloc à la position (n,m) .

$DC(I_t)$ est l'expression du contenu global de l'image. Il est calculé par

5 une moyenne des activités $(bsa_{n,m})$ concernés.

$$DC(I_t) = \overline{dc_{n,m}(I_t)} = \frac{1}{N \times M} \sum_{n=0}^{H-1} \sum_{m=0}^{W-1} dc_{n,m} \quad (9)$$

La figure 7 illustre les variation du paramètre DC pour la même séquence que la figure 6. L'allure de la courbe est sensiblement la même que dans le cas de la figure 6, c'est-à-dire qu'on distingue la trace des différents évènements intervenus au cours de la séquence.

A partir d'un paramètre d'activité tel que $bsa_{n,m}$, on peut définir un contraste bloc local $c_{n,m}$:

$$c_{n,m} = \frac{\sqrt[p]{bsa_{n,m}}}{dc_{n,m}} \quad (10)$$

15 p étant une constante (par exemple p entier ≥ 2).

Il en découle un contraste bloc moyen CBM .

Pour l'image I_t , ce contraste CBM (I_t) est définie comme :

$$CBM(I_t) = \overline{c_{n,m}(I_t)} = \frac{1}{N \times M} \sum_{n=0}^{H-1} \sum_{m=0}^{W-1} c_{n,m}(I_t) \quad (11)$$

20 Un contraste global de l'image $CG(I_t)$ peut être défini sur le même principe, mais à partir des paramètres globaux $DC(I_t)$ et $SA(I_t)$.

$$CG(I_t) = \frac{\sqrt{SA(I_t)}}{DC(I_t)} \quad (12)$$

L'évolution du paramètre CG pour la même séquence que celle utilisée précédemment pour les figures 6 et 7 est représentée à la figure 8. On voit que seul le dépointage intervenu à l'image 260 entraîne une modification significative de la valeur de CG.

5 Les paramètres de contraste sont donc sensibles au contenu de l'image et seulement à celui-ci, indépendamment des variations de luminosité.

Pour définir un critère de détection d'un changement de contenu, le principe est de vérifier que soit l'activité spatiale, soit le contraste
10 varient peu d'une image à l'autre, en s'affranchissant des variations de luminosité.

Dans l'un et l'autre cas, ceci revient à une pondération par un facteur représentatif de la luminosité.

Dans le premier cas, le critère de détection d'un changement
15 de contenu basé sur l'activité spatiale bloc $bsa_{(n,m)}$ correspondant au bloc (n,m) doit s'affranchir au maximum des variations de luminosité, car la valeur de l'activité spatiale dépend de la luminosité.

Pour détecter un changement de contenu, on va examiner l'évolution de l'activité spatiale $bsa_{n,m}(t)$, caractérisant un bloc d'une image
20 courante, à la date t , par rapport à une image antérieure $l(t')$, qui est de préférence l'image immédiatement précédente $l(t-1)$. Le rapport $k_{n,m}(t)$ des deux valeurs est utilisé :

$$k_{n,m}(t) = bsa_{n,m}(t) / bsa_{n,m}(t') \quad (13)$$

25 Pour définir une signature caractéristique des éléments contenus dans l'image, et en principe indépendante de la luminosité, il faut s'efforcer de compenser les effets des variations de luminosité sur l'activité spatiale bloc.

L'activité spatiale diminue avec l'éclairement, malgré le
30 dispositif d'adaptation automatique de la caméra, d'un facteur $k_{n,m}$ propre à chaque bloc. Cependant, une variation de luminosité est globale, c'est-à-dire que chaque bloc de l'image est, normalement, affecté de la même manière. On peut alors estimer qu'en cas de variation de l'éclairement, le facteur de changement d'activité spatiale $k_{n,m}$ d'un bloc, est le même que la variation du

facteur $K(t)$ de changement d'activité spatiale moyenne SA entre l'image $I(t)$ et l'image $I(t')$ à l'instant t avec :

$$K(t) = SA(t) / SA(t').$$

L'activité spatiale moyenne SA est définie ci-dessus.

5 Pour les blocs n,m dont le facteur $k_{n,m}$ est similaire à $K(t)$, on pourra dire que leur contenu a peu changé d'une image à l'autre, malgré la prise en compte des effets de variation de luminosité. Inversement, pour les blocs dont le contenu a changé, les valeurs $k_{n,m}(t)$ et $K(t)$ seront très différentes. Le critère $\text{caract}_{n,m}$, qui est utilisé de manière préférentielle pour
10 caractériser le changement de contenu indépendamment de la luminosité est défini de la manière suivante :

On a :

$$\text{ratio}_{n,m}(t) = k_{n,m}(t) / K(t) = [bsa_{n,m}(t) / bsa_{n,m}(t')] / [SA(t) / SA(t')] \quad (14)$$

Le critère $\text{caract}_{n,m}$ est défini comme la valeur maximale

$$\text{caract}(t)_{n,m} = \max[\text{ratio}_{n,m}(t), 1 / \text{ratio}_{n,m}(t')] - 1 \quad (15)$$

15 Il s'agit de la valeur maximale du rapport entre $\text{ratio}_{n,m}(t)$ et $\text{ratio}_{n,m}(t')$ à laquelle on retranche le chiffre 1, c'est-à-dire :

si $\text{ratio}_{n,m}(t) > \text{ratio}_{n,m}(t')$, alors $\text{caract}_{n,m}(t) = (\text{ratio}(t) / \text{ratio}(t')) - 1$

si $\text{ratio}_{n,m}(t) < \text{ratio}_{n,m}(t')$, alors $\text{caract}_{n,m}(t) = (\text{ratio}(t') / \text{ratio}(t)) - 1$

si $\text{ratio}_{n,m}(t) = \text{ratio}_{n,m}(t')$, alors $\text{caract}_{n,m}(t) = 0$

20 De la sorte, cette valeur est toujours supérieure ou égale à 0.

Le critère $\text{caract}_{n,m}$ décide qu'un bloc de l'image change de contenu entre deux images de type « / » dont l'une est une image courante $I(t)$ et l'autre une image antérieure $I(t')$ lorsque la variation d'activité bloc est très différente de la variation d'activité image (définie comme l'activité spatiale
25 moyenne des de l'image). La valeur de $\text{caract}_{n,m}$ est toujours positive ou nulle, que la luminosité augmente ou diminue. Cette propriété permet de n'utiliser qu'un seuil de détection au lieu de deux. On considère qu'un bloc a changé de contenu lorsque cette caractéristique a atteint une certaine valeur
seuilVariationContenu :

30

$$caract_{n,m} > \text{seuilVariation Contenu}$$

On peut utiliser bien entendu un autre critère tenant compte
 5 de $\text{ratio}(t)$ et de $\text{ratio}(t')$. On peut par exemple utiliser $\text{ratio}(t)/\text{ratio}(t')$ et mettre
 en oeuvre deux seuils, qui constituent deux seuils d'intrusion prédéterminés.

Une illustration de l'efficacité de ce critère est donnée ci-dessous.

Si on calcule le paramètre $caract_{n,m}(t)$ de la comparaison de
 10 l'activité spatiale bloc avec compensation des effets de variation de
 luminosité, entre deux images, l'une prise en éclairage artificiel et l'autre après
 extinction de cet éclairage et réadaptation, on peut constater que l'utilisation
 de la compensation des effets de variation de luminosité permet de limiter les
 écarts non liés à un changement de contenu. Certains blocs ne sont
 15 cependant pas affectés par la variation de la luminosité, notamment ceux du
 contour des moniteurs informatiques allumés.

Dans le deuxième cas, l'utilisation d'un paramètre basé sur le
 contraste pour définir un critère de détection de variation de contenu est plus
 simple. En effet, le contraste est indépendant de la luminosité, donc la
 20 comparaison peut être faite directement :

$$k'_{n,m}(t) = c_{n,m}(t) / c_{n,m}(t') \quad (17)$$

On utilise préférentiellement un critère $caract'_{n,m}$ défini comme :

$$caract'_{n,m} = \max[k_{n,m}(t), 1 / k_{n,m}(t')] - 1 \quad (18)$$

Il s'agit de la valeur maximale du rapport entre les deux valeur
 $k_{n,m}(t)$ et le $k_{n,m}(t')$ à laquelle on retranche 1, de sorte que ce rapport est
 25 toujours supérieur ou égal à 0.

Le critère $caract_{n,m}$ détecte une variation de contenu d'un bloc
 si la condition suivante est réalisée :

$$caract'_{n,m} > \text{seuilVariation Contenu} \quad (18')$$

En général, pour calculer le paramètre $\text{caract}_{n,m}$ ou $\text{caract}'_{n,m}$, on choisit l'image courante (t) à l'instant t et l'image précédente l(t-1) à l'instant t' = t-1. On peut également choisir une autre image antérieure l. Lorsque le seuil défini en (18') est dépassé, on cesse de mettre à jour l'image précédente qui sert de base à la comparaison.

L'organigramme de la phase d'apprentissage est représenté sur la figure 9. Il est appliqué pour chaque nouvelle image présentée à l'entrée. Elle met en œuvre trois étapes : une présélection des blocs de l'image, l'élimination des blocs instables, et enfin la mise à jour du dictionnaire D par l'élimination des blocs instables repérés.

Après le traitement de chaque image, un test détermine si l'apprentissage est terminé, puis si le dictionnaire résultant D peut être utilisé par la phase d'exploitation.

Durant la phase d'apprentissage, les blocs témoins sont sélectionnés et leur emplacement dans l'image est mémorisé. Les blocs témoins sont ceux qui contiennent des éléments caractéristiques et dont le contenu ne change pas au sens de l'un ou l'autre critère $\text{caract}_{n,m}$ ou $\text{caract}'_{n,m}$ définis ci-dessus.

L'ensemble des blocs témoins forme un dictionnaire de référence D.

Pour cela, les changements intervenant entre deux images séparées d'une durée définie en fonction des besoins (par exemple 1 seconde) sont observés bloc par bloc. Tout bloc détecté une fois par le critère $\text{caract}_{n,m}$ ou par le critère $\text{caract}'_{n,m}$ est éliminé du dictionnaire D des blocs témoins.

Les blocs témoins peuvent être choisis comme présentant un contenu remarquable. Cette caractéristique est mise en évidence lors de l'étape de présélection des blocs témoins. Celle-ci se déroule une seule fois, au début de la phase d'apprentissage. Cette étape consiste à sélectionner les blocs de contenu structuré, contenant en particulier des éléments horizontaux ou verticaux nets. Ceux-ci sont en effet courants dans les images (mobilier, murs, embrasure de porte, etc...). Toutefois, suivant la scène observée, d'autres structures caractéristiques peuvent être choisies.

Pour les éléments verticaux et horizontaux, la solution proposée est la suivante (voir l'organigramme de la figure 10).

On applique séparément à l'image pixel f les filtres Sobel de masque $[1-1]$ (horizontal) et $[1-1]^t$ (vertical). Puis, l'activité de l'image pixel résultante dont le contour a été ainsi isolé par les filtres de Sobel est calculée pour chaque bloc $B_{n,m}$ selon la formule (5) ou (7). On obtient donc deux images de l'activité spatiale bloc orientée, respectivement horizontale
 5 $asbo_H(n,m)$ et verticale $asbo_V(n,m)$, et d'activité moyenne orientée ASo_H et ASo_V calculée selon (6).

Pour sélectionner les blocs d'activité orientée suffisante, on sélectionne l'activité orientée la plus grande (parmi l'activité orientée
 10 horizontale et verticale), et on impose que cette dernière soit supérieure à un seuil (FS) représentant un rapport entre l'activité orientée du bloc considéré et l'activité spatiale orientée moyenne de l'image.

On note FS_H le seuil retenu pour l'activité orientée verticalement, et FS_V pour l'activité orientée horizontalement :

15

si $(asbo_H(n,m) > asbo_V(n,m))$

et si $(asbo_H(n,m) > FS_H)$, alors on sélectionne le bloc

sinon

si $(asbo_V(n,m) > FS_V)$, alors sélectionner le bloc (19)

Les blocs sélectionnés à l'issue de ce traitement forment un premier dictionnaire qui va être affiné sur une longue période et de nombreuses images par les étapes suivantes.

20 Les blocs témoins sont également caractérisés par un contenu invariant entre deux images. Un critère de détection d'un changement de contenu d'un bloc, basé sur l'activité spatiale du bloc, a été défini précédemment. Ce critère est appliqué ici : tout emplacement de l'image où un changement de contenu est détecté dans un bloc est éliminé de
 25 l'ensemble des blocs témoins.

Cette étape est itérée sur de nombreuses images et sur une durée suffisante pour placer le système dans une gamme de conditions de fonctionnement large, représentative des conditions qui seront normalement rencontrées durant une phase d'exploitation. Par exemple, la phase
 30 d'apprentissage peut inclure des variations de luminosité (telles que celles du cycle jour/nuit), ou d'autres perturbations telles que le passage d'objets dans

une partie du champ de la caméra, ou une intrusion dans une zone de passage, l'allumage ou l'extinction d'une lumière ou d'un écran d'ordinateur etc ...

Une autre tâche importante est réalisée à cette étape : il s'agit également de mettre à jour l'image qui sera comparée à la prochaine image qui se présentera. Pour cela, on mémorise simplement l'image courante, car aucune alarme n'est supposée se produire durant la phase d'apprentissage.

La phase d'apprentissage prend fin lorsque la durée prévue est expirée. La durée de la phase d'apprentissage doit couvrir une période suffisamment longue pour inclure le maximum de phénomènes perturbateurs. Cependant, afin de raccourcir ce délai, on peut également stopper la phase d'apprentissage si le dictionnaire obtenu n'a pas varié depuis une longue durée (par exemple plusieurs heures).

La condition stoppant l'apprentissage est donc :

15

$$Durée\ apprentissage \geq duréeMaxiPhaseApprentissage$$

ou

$$DuréeDicoInvariant \geq duréeMaxiSansChangement \quad (20)$$

Enfin, il reste à vérifier que la phase d'apprentissage a réussi, c'est-à-dire que le dictionnaire est valide. Pour cela, il suffit qu'il atteigne une taille suffisante, ou autrement dit qu'il y ait dans l'image un nombre significatif de blocs témoins. Cette condition permet notamment d'éviter de déclencher aléatoirement des alarmes lors de la phase d'exploitation : un exemple typique est celui d'une scène dans laquelle la zone de passage des objets concerne la quasi-totalité de l'image. Cette situation aboutit à un nombre très réduit de blocs stables. Lors de la phase d'exploitation ou de détection, le changement de contenu de quelque blocs suffira à déclencher une fausse alarme.

25

La condition validant le dictionnaire obtenue est :

$$Taille(Dictionnaire) > proportionBlocsDansDico \times (H \times W), \quad (21)$$

où $(H \times W)$ est le nombre de blocs de l'image.

Un module peut détecter les situations particulières d'obstruction totale. Il fonctionne de manière totalement indépendante du

30

détecteur de dépointage / obstruction partielle. Afin de différencier le cas d'une obstruction totale de celui d'une scène entièrement noire ou uniforme, on considère que :

- une obstruction totale du champ d'une caméra est un phénomène quasi instantané (l'image passe d'une image ayant un contenu à une image noire de façon instantanée) ;
- une image noire provenant d'une scène non éclairée est obtenue progressivement.

C'est sur ce postulat qu'est basée la détection d'une obstruction totale. Si l'on passe de façon brutale d'une image contenant de l'information à une image noire, on considère qu'il y a obstruction totale. Par contre, si le changement est progressif, on ne considère pas l'événement comme une obstruction totale.

Dans un premier temps, on synthétise les évolutions passées en calculant la moyenne $SA_{moy}(t)$ de l'activité spatiale image $SA(t)$, sur les MEM_IMG_AS images arrivées précédemment à l'image courante, et sur lesquelles aucune obstruction totale n'a été détectée.

$$SA_{moy}(t) = \frac{1}{MEM_IMG_AS} \sum_{\tau=1}^{MEM_IMG_AS} SA(t-\tau) \quad (22)$$

20

Une baisse de luminosité de l'image est interprétée comme une obstruction totale si on a à la fois une forte baisse de l'activité spatiale $SA(t)$ par rapport à l'activité spatiale moyenne $SA_{moy}(t)$, une valeur faible de l'activité spatiale $SA(t)$, et une valeur significative de l'activité spatiale moyenne passée $SA_{moy}(t)$, avec par exemple ::

25

$$SA(t) \times 2 < SA_{moy}(t)$$

$$\text{et } SA(t) < \text{seuilActiBlocMinimale} \quad (23)$$

$$\text{et } SA_{moy}(t) > \text{SeuilActiBlocMinimale}$$

Un fonctionnement du détecteur de dépointage / obstruction peut être obtenu dans des conditions limites, de la manière suivante :

Un module a pour but d'éliminer les situations dans lesquelles l'image analysée est trop peu « riche », et en particulier trop sombre, pour prendre toute décision fiable de déclenchement d'alarme. Pour cela, on utilise une condition de luminosité d'image, qui se traduit par une activité spatiale de l'image suffisante. Lorsque la luminosité baisse en dessous du seuil acceptable, par exemple à la tombée de la nuit, le détecteur suspend toute surveillance de la variation du contenu des blocs témoin, et se met en veille. Lorsque les conditions sont redevenues favorables (éclairage artificiel ou lever du jour), le détecteur se réactive automatiquement et compare le contenu de la nouvelle image à celui de la dernière image exploitable avant la mise en veille. Ainsi, tout dépointage survenu durant la période d'obscurité sera détecté.

15 $SA(t) > \text{seuilActiBlocMinimale}$

(24)

Lorsque la condition (24) n'est pas satisfaite, l'alarme de dépointage / obstruction est éventuellement ramenée à FAUX, et un indicateur « fonctionnement impossible » positionné à VRAI. Typiquement, ce cas se produit lorsque la scène observée n'est pas éclairée la nuit. Le détecteur attend l'image suivante pour éventuellement réactiver le détecteur.

20 La phase d'exploitation consiste à surveiller les blocs témoins désignés par le dictionnaire de référence issu de l'apprentissage. Pour cela, on applique le même critère de détection de changement de contenu entre deux images « I » successives que lors de l'apprentissage.

25 Le procédé compte le nombre de blocs témoins du dictionnaire, ayant changé de contenu. Si ce nombre est significatif, on détecte un dépointage / obstruction de la caméra. Cependant, une instabilité des blocs du dictionnaire de référence peut se produire lorsqu'un changement brusque de luminosité n'a pas laissé le temps à la caméra de rétablir une image nominale. Par conséquent, afin de limiter les fausses alarmes, un événement de dépointage / obstruction ne sera réellement signalé que s'il est détecté sur au moins deux fois consécutivement (ou plus).

30 L'organigramme de la phase d'exploitation est représenté sur la figure 11. Comme pour la phase d'apprentissage, il est appliqué pour chaque image présentée à l'entrée.

On retrouve également les modules identiques d'activation ou désactivation du détecteur, et de détection de l'obstruction totale.

Les deux autres étapes ont de nombreux points communs avec la phase d'apprentissage.

5 La première de ces deux étapes est le repérage des blocs du dictionnaire D issue de l'apprentissage qui sont instables dans la scène observée. Le processus de déclenchement de l'alarme de dépointage/obstruction constitue la seconde.

10 La détection des blocs de contenu variable est réalisée en calculant le paramètre caractéristique du changement de contenu $\text{caract}_{n,m}$ ou $\text{caract}'_{n,m}$. Le processus est identique à celui de la phase d'apprentissage.

Durant la phase d'apprentissage, l'algorithme s'est attaché à conserver seulement les blocs $B_{n,m}$ qui n'ont pas changé de contenu d'une image à l'autre, selon le critère $\text{caract}_{n,m}$ ou $\text{caract}'_{n,m}$. Ces blocs témoins sont
15 caractéristiques de la scène vis-à-vis des tentatives de dépointage ou d'obstruction.

Durant la phase d'exploitation, l'algorithme compte le nombre de blocs témoins ayant changé de contenu entre deux images qui sont soit l'image courante et une image précédente, notamment l'image
20 immédiatement précédente, soit l'image courante et une image de référence qui est éventuellement rafraîchie périodiquement. Si ce nombre est important, on détecte un dépointage/obstruction de la caméra :

$\text{nbreBlocsDicoDetectes} > \text{proportionBlocsAlarm} \times \text{Taille(Dictionnaire)}$,

Si cette condition est réalisée, l'alarme est déclenchée.

25 Ce module réalise également la mise à jour de l'image qui sera comparée à la prochaine image qui se présentera. Pour cela et selon un mode préféré, on mémorise simplement l'image courante $I(t)$, mais seulement à la condition qu'un événement de dépointage/obstruction n'y ait pas été détecté, et cette image courante servira de base de comparaison à la
30 prochaine image courante qui sera captée.

Un exemple de mise en œuvre du procédé est donné en figure 12. Le Processeur de calcul du Critère de détection de Changement de Contenu (PCCC) fournit le paramètre $\text{caract}_{n,m}(t)$ ou $\text{caract}'_{n,m}(t)$. Ce paramètre est calculé comme décrit ci-dessus, en se basant exclusivement
35 sur l'activité spatiale $\text{bsa}_{n,m}(t)$ et sur des blocs de taille égale à la taille couramment utilisée pour le codage des images, c'est-à-dire $M = 8$.

Durant la phase d'apprentissage, le Processeur de Détection de Dépointage/Obstruction partielle (PDDO) construit progressivement le dictionnaire D en éliminant les blocs ayant changé de contenu et ne pouvant donc faire partie des blocs témoins. Lors de la phase d'exploitation le PDDO
5 compte les blocs de ce dictionnaire ayant changé de contenu, pour éventuellement déclencher une alarme.

Les processeurs PDIU et PDOT fonctionnent en parallèle des processeurs PCCC et PDDO. Le processeur PDIU indique au processeur PDDO si l'image est inutilisable, et dans ce cas, celui-ci désactive l'indicateur
10 de dépointage et/ou obstruction partielle.

Exemple

Pour une scène prise en milieu intérieur, pouvant être éclairée artificiellement, et dans laquelle des personnes peuvent rentrer dans le champ de la caméra, des images vidéo 256 x 256 ont été traitées par blocs de tailles
15 8 x 8, soit en tout 1024 blocs.

Après une phase de détection des blocs orientés, le dictionnaire comporte 613 blocs (voir figure 13).

Au cours de la phase d'apprentissage pendant de laquelle les blocs non stables sont éliminés, le dictionnaire D se stabilise autour de 400
20 blocs (411 blocs exactement).

La grande majorité des blocs se trouvant dans une zone de passage ont été éliminés.

En dehors de la zone de passage, d'autres blocs ont été éliminés : ce sont ceux qui sont devenus instables lors d'un changement d'éclairage intervenu au cours de l'apprentissage.
25

Les blocs sélectionnés, conservés dans le dictionnaire D, perdent leur stabilité en cas de dépointage et/ou d'une obstruction, ce qui permet une détection spécifique.

La figure 14 illustre le comportement de détecteur. Lorsqu'aucun événement ne vient perturber l'image, le nombre de blocs pour
30 lesquels un changement d'activité est constaté est faible ou nul (entre 0 et 2).

A l'image n° 14, un premier pic (27 blocs) en I correspondant au passage d'une personne est détecté. Cette détection n'est pas prise en compte pour déclencher l'alarme.

A l'image n° 30, une augmentation II, puis une stabilisation III du nombre de blocs actifs détectés apparaît. Elle est due à un début de dépointage de la caméra.

5 A l'image n° 38, un pic important IV du nombre de blocs actifs (100 blocs) correspond à un dépointage franc et au déclenchement de l'alarme de dépointage.

On compte donc que, lors d'un dépointage, le nombre de blocs actifs (pour lesquels l'activité est détectée par l'évolution du critère caract ou caract') devient très important, ce qui permet une détection
10 spécifique d'un dépointage, quelles que soient les variations d'éclairage.

La figure 15 illustre le cas d'une obstruction totale de la caméra. L'image passe d'une image ayant un contenu à une image entièrement noire (image n° 23). On détecte une chute brutale V de l'activité moyenne AS de l'image dès que l'objectif est obstrué (image n° 2).

15 En cas de diminution progressive de la luminosité d'une scène filmée (par passage progressif de la lumière du jour à la nuit), la chute de l'activité moyenne de l'image n'est pas brutale, mais au contraire très progressive. Dans un tel cas, il n'y a pas déclenchement d'une alarme.

REVENDECATIONS

1. Procédé de détection d'intrusion dans un système de télésurveillance mettant en œuvre au moins une caméra vidéo fournissant une image vidéo, caractérisé en ce qu'il met en œuvre pour une image résultante obtenue par soustraction entre les valeurs pixel par pixel d'une image courante et d'une image de référence, une transformée par blocs, dans au moins une zone caractéristique de l'image comprenant N blocs, pour obtenir pour chaque bloc B_n , m des coefficients transformés $F_{n,m}(i,j)$ i et j variant de 0 à M, M désignant le nombre maximal de lignes ou de colonnes de pixels d'un bloc $B_{n,m}$, puis le calcul, à partir desdits coefficients $F_{n,m}(i,j)$ obtenus, d'une activité spatiale moyenne A_Z des N blocs de ladite zone de l'image résultante, qui constitue un signal numérique de détection d'intrusion qui est comparé avec un seuil prédéterminé de détection d'intrusion pour détecter une intrusion.

2. Procédé selon la revendication 1, caractérisé en ce qu'il met en œuvre, pour le calcul de l'activité spatiale moyenne A_Z :

- le calcul de l'activité spatiale A_{Su} de chacun des N blocs B_n , m, avec u variant de 0 à N - 1.
- Le calcul de l'activité spatiale moyenne A_Z , avec :

$$A_Z = \frac{1}{N} \sum_{u=0}^{N-1} A_{Su}$$

3. Procédé selon la revendication 1, caractérisé en ce qu'il met en œuvre pour le calcul de l'activité spatiale moyenne A_Z :

- le calcul de l'activité spatiale A_{Su} de chacun des N blocs B_n , m, u variant de 0 à N-1,
- le calcul de l'activité spatiale moyenne A_Z , avec :

$$A_Z = \frac{1}{N} \sum_{u=0}^{N-1} C_{Su}, \text{ avec } C_{Su} = \frac{\sqrt{A_{Su}}}{d_{Cn}}$$

avec p constante (par exemple $p \geq 2$),

d_{Cu} représentant la luminosité moyenne du bloc de rang u .

4. Procédé selon une des revendications 1 à 3, caractérisé en ce qu'il met en œuvre l'application d'une transformée temporelle unidimensionnelle à N' signaux de détection d'intrusion ($A_{Z0} \dots A_{ZN'-1}$) calculés pour des images courantes prises à des instants $t = 0 \dots t = N' - 1$, pour obtenir une matrice de coefficients transformés $D_0, D_1 \dots D_{N'-1}$ et un calcul d'un signal A_T représentatif d'une évolution temporelle de l'image avec :

$$A_T = \left(\sum_{u=1}^{N'-1} D_u^2 \right)$$

5. Procédé selon la revendication 4, caractérisé en ce qu'il met en œuvre la comparaison du signal d'évolution temporelle A_T avec un seuil et déclenchement d'une action de détection d'intrusion si ledit signe A_T franchit ledit seuil.

6. Procédé selon une des revendications précédentes, caractérisé en ce qu'il met en œuvre plusieurs signaux de détection (A_Z) correspondant à des zones caractéristiques distinctes de l'image.

7. Procédé selon une des revendications précédentes, caractérisé en ce qu'il met en œuvre au moins un signal de détection (A_Z) élaboré à partir d'au moins deux zones caractéristiques de l'image.

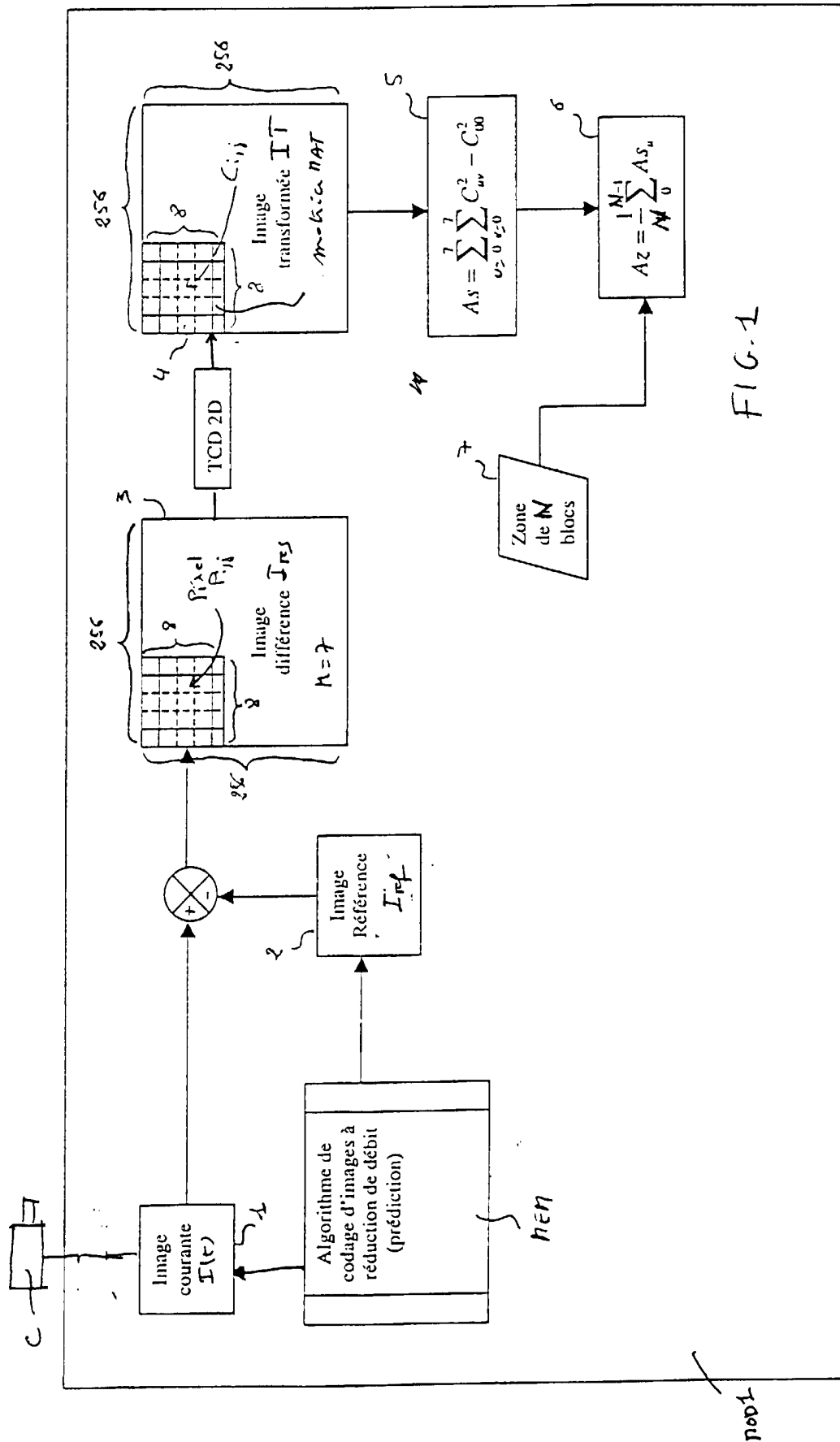
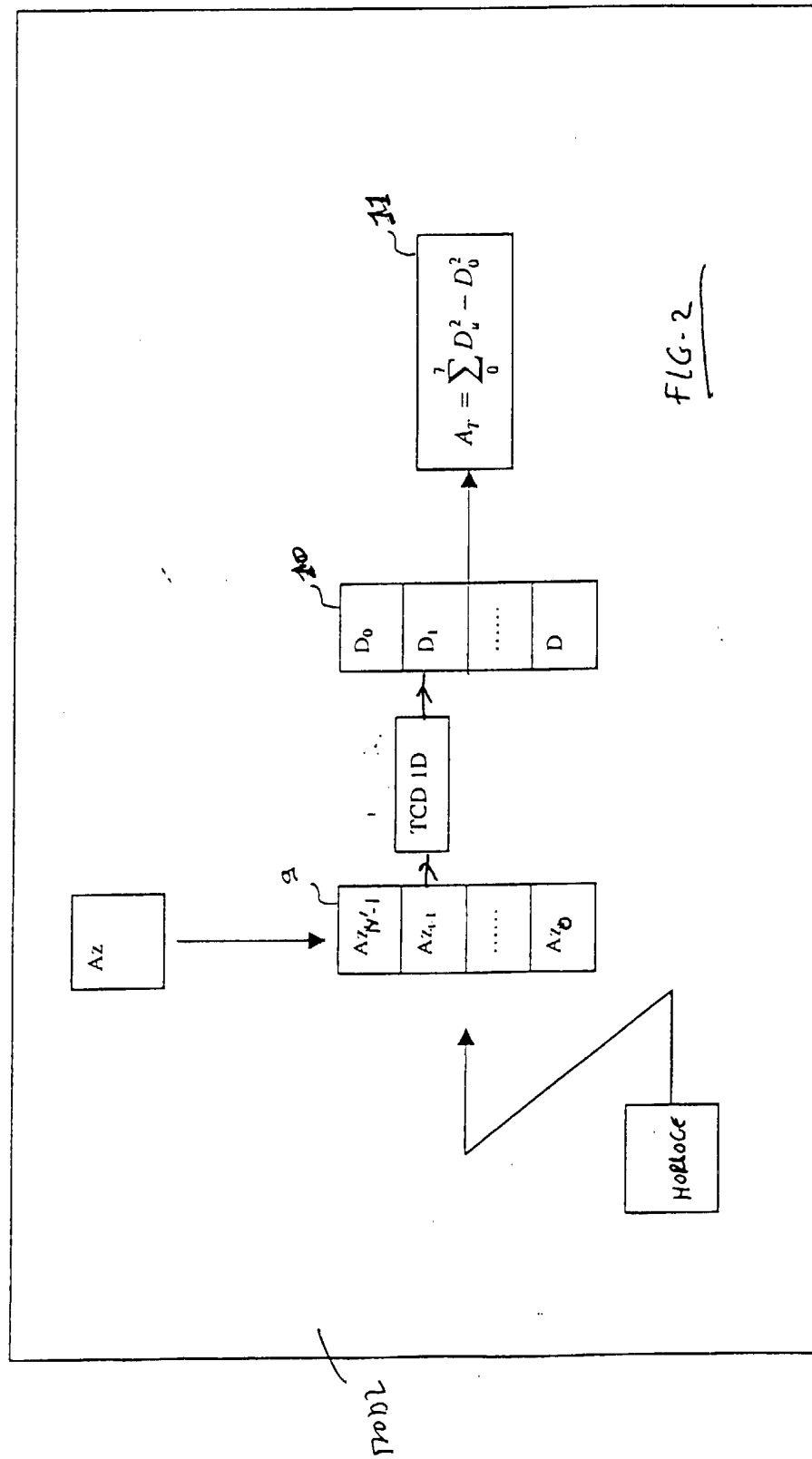
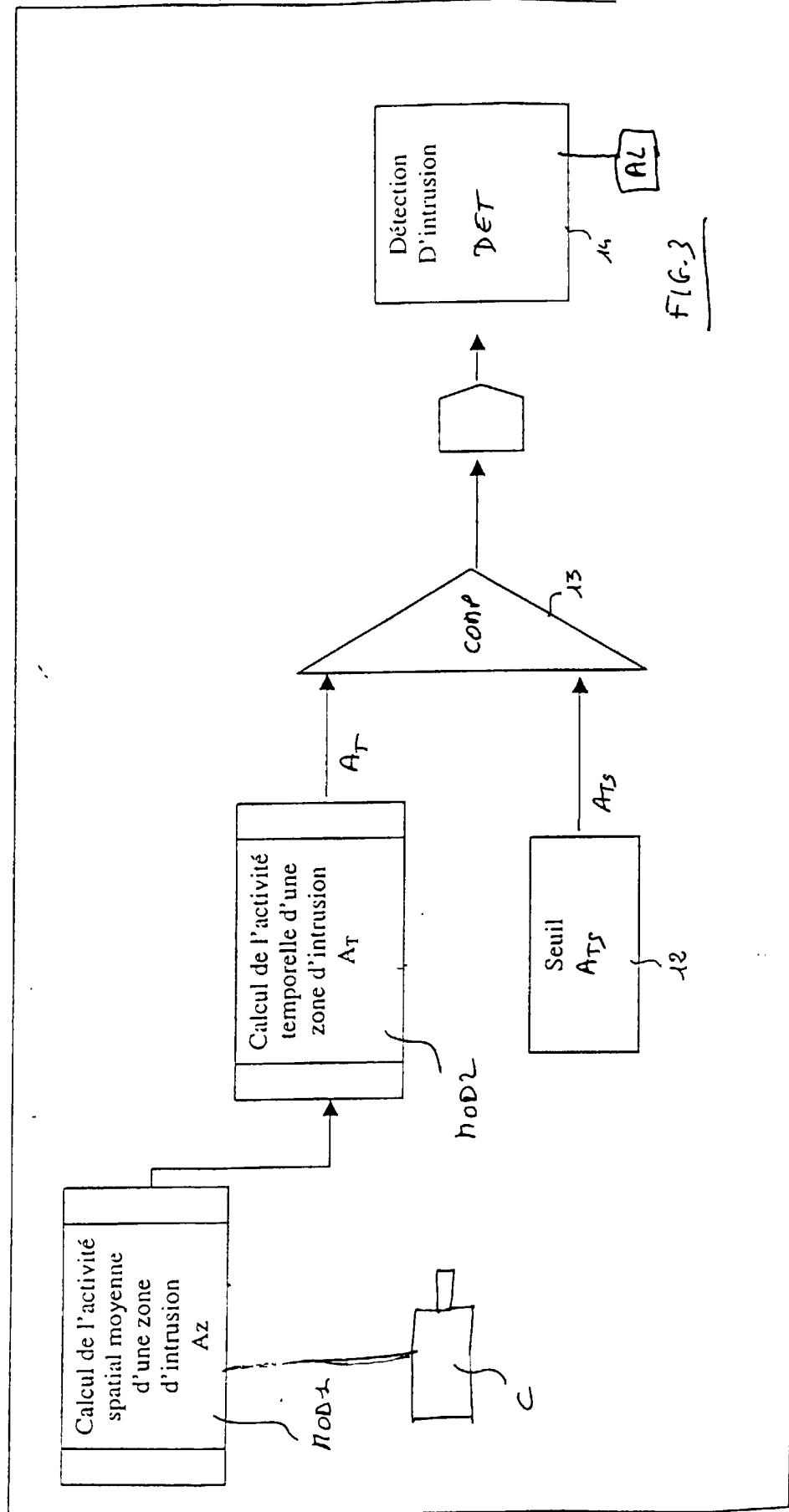


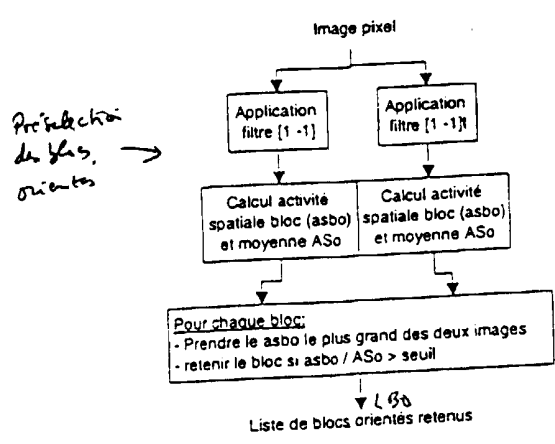
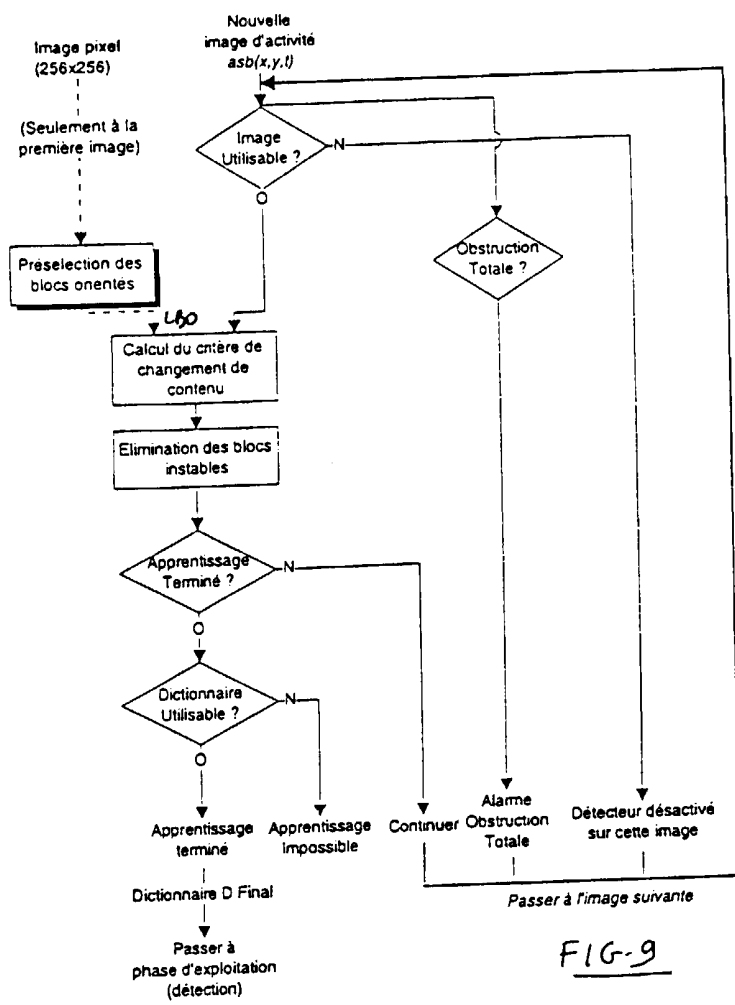
FIG. 1





PAS DE FIGURE 4, NI DE FIGURES 5, 6, 7, 8

5/7



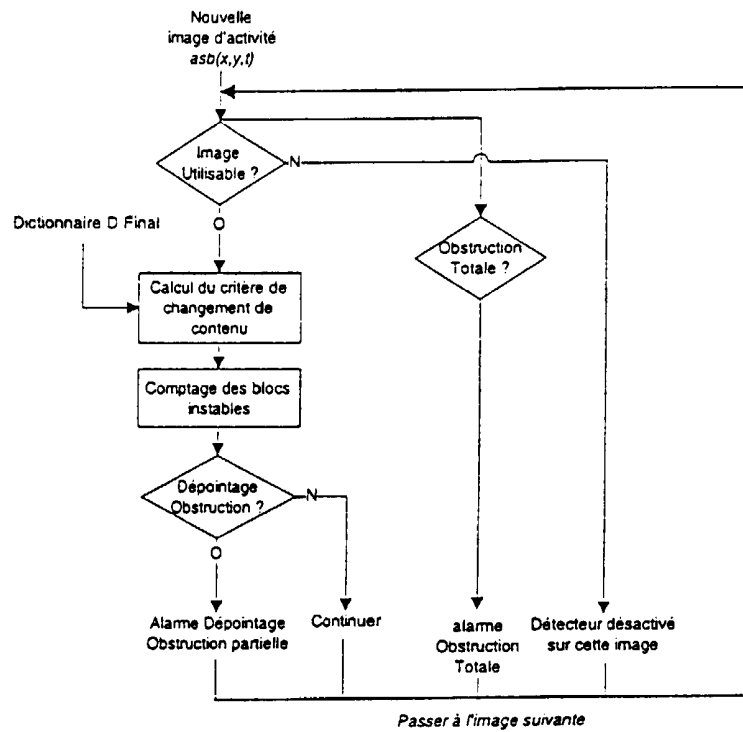


FIG. 11

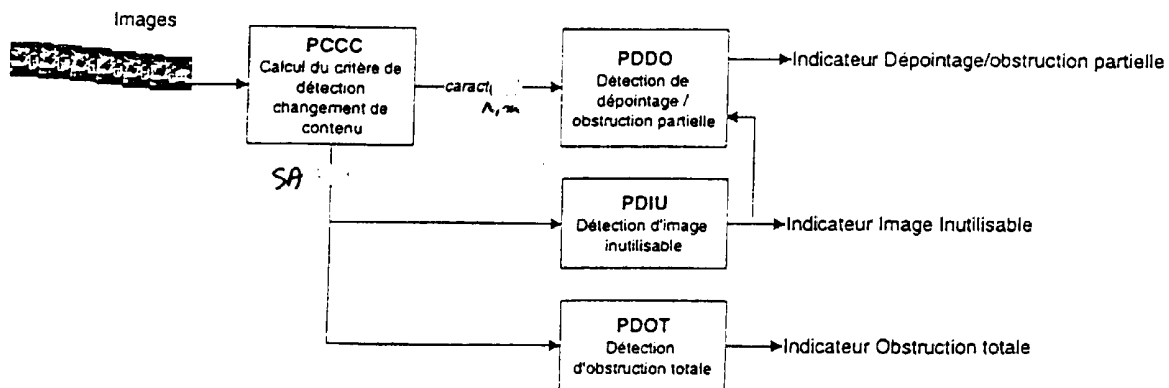


FIG. 12

7/7

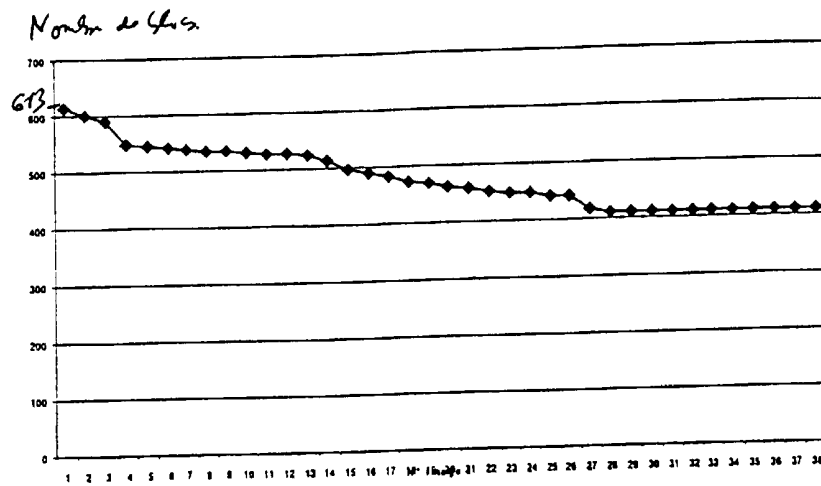


FIG 13

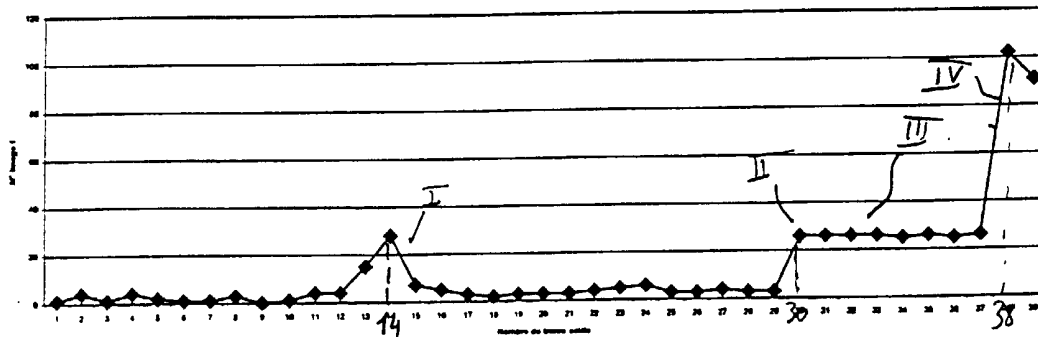


FIG 14

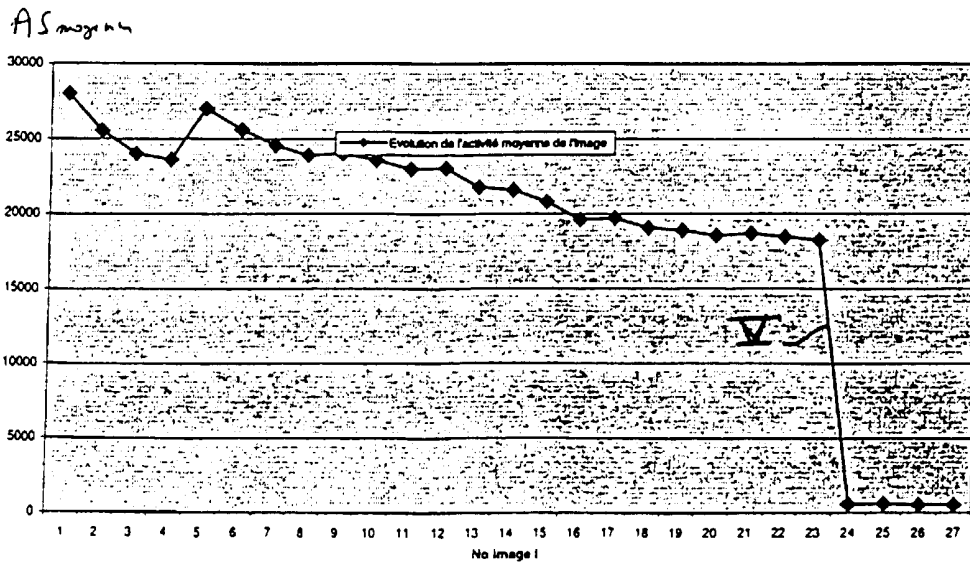


FIG. 15