

DESCRIPTION**Titre de l'invention : DISPOSITIF ET PROCEDE D'AIDE A LA DECISION POUR LA GESTION DE CONFLITS AERIENS****Domaine technique**

L'invention concerne de manière générale les systèmes d'aide à la décision, et en particulier un système et un procédé d'aide à la décision pour la gestion de conflits aériens.

Art Antérieur

[0001] Le développement de systèmes d'aide à la décision a connu un essor croissant ces dernières années et s'est étendu à de nombreux secteurs industriels, notamment dans les secteurs où il existe un enjeu de sécurité, comme par exemple dans le domaine des systèmes de contrôle aérien.

[0002] Les systèmes de contrôle aérien doivent assurer la sécurité du trafic aérien. Les systèmes de contrôle aérien sont conçus pour garantir des distances de sécurité entre les avions de leurs secteurs tout en maintenant des distances minimales de sécurité entre les avions dont les trajectoires vont se rapprocher, en modifiant au moins une de ces trajectoires. Les systèmes de contrôle aérien connus sont équipés d'outils de contrôle du trafic aérien qui permettent notamment la détection de conflits aériens, et/ou fournissent une aide à la décision pour la gestion des conflits aériens.

[0003] Il existe deux approches connues pour la gestion des conflits aériens.

[0004] Une première approche se base sur des calculs géométriques pour assurer une décision continue dans le temps, ce qui implique une utilisation intense de ressources de calcul puissantes.

[0005] Une seconde approche repose sur l'utilisation d'algorithmes d'intelligence artificielle pour la résolution de conflits aériens tout en minimisant les ressources requises pour les calculs.

[0006] Par exemple, dans l'article « Reinforcement Learning for Two-Aircraft Conflict Resolution in the Presence of Uncertainty, Pham *et al.*, Air Traffic Management Research Institute, School of Mechanical and Aerospace Engineering Nanyang Technological University, Singapore, March 2019», une solution d'apprentissage par renforcement est proposée pour automatiser la résolution de conflits aériens. Une telle solution est prévue pour assurer le maintien de distances de séparation minimales. Cette solution met en œuvre un algorithme appelé 'Deep Deterministic Policy Gradient' utilisant un réseau de neurones artificiels dense qui permet une résolution de conflit restreinte à deux avions ne naviguant qu'en ligne droite et à un espace bidimensionnel avec une seule action possible de résolution

qui consiste en un changement de direction avec retour sur un point nommé de la trajectoire initiale.

[0007] L'article « Autonomous Air Traffic Controller : A Deep Multi-Agent Reinforcement Learning Approach, Marc Brittain, Peng Wei, Department of Aerospace Engineering, Iowa State University, May 2019 » décrit une autre solution d'apprentissage par renforcement pour automatiser la résolution de conflits aériens, la solution mettant en œuvre un algorithme par renforcement multi-agent profond ('Deep multi-agent reinforcement learning' en langue anglo-saxonne) avec réseaux de neurones artificiels denses pour l'approximation. Cette solution permet la résolution de conflits sans restriction sur le nombre d'avions. Cependant, la résolution de conflits aériens selon cette solution est restreinte à un espace bidimensionnel avec pour seule action possible de résolution le changement de vitesse. En outre, le réseau de neurones mis en œuvre dans cette solution doit être ré-entraîné pour chaque type de scénario et ne permet pas la généralisation à une nouvelle suite de points nommés.

[0008] L'article « Autonomous Aircraft Sequencing and Separation with Hierarchical Deep Reinforcement Learning, Marc Brittain, Peng Wei, Department of Aerospace Engineering, Iowa State University, 2018 » décrit également une solution d'apprentissage par renforcement pour la résolution de conflits aériens. Cette solution permet un choix des plans de vols en utilisant deux réseaux de neurones imbriqués l'un dans l'autre de manière à ce qu'un premier réseau (« réseau père ») choisisse les plans de vol, et que le second réseau (« réseau fils ») régule la vitesse pour maintenir la séparation entre les avions. Cette solution permet le maintien de la séparation et la résolution du conflit si la séparation est perdue ainsi qu'une minimisation du temps de trajet. Cependant, la résolution de conflit selon cette solution est restreinte à un espace bidimensionnel avec pour seule action possible de résolution le changement de vitesse. De plus, cette solution fonctionne pour un nombre très limité d'avions et requiert l'entraînement des réseaux de neurones pour chaque type de scénario.

[0009] Les solutions existantes de gestion des conflits aériens se limitent toutefois à un nombre restreint de configurations possibles en termes de nombre d'avions, de couloirs aériens, de catégories d'avions, de vitesses ou d'altitudes d'avions, ou encore d'actions possibles pour résoudre les conflits détectés.

[0010] Il existe donc un besoin pour un système et un procédé de gestion du trafic aérien amélioré capable de résoudre efficacement les conflits aériens.

Définition Générale de l'invention

[0011] L'invention vient améliorer la situation. A cet effet, l'invention propose un dispositif de gestion de trafic aérien, dans un espace aérien comprenant un aéronef de référence et au moins un autre aéronef, le dispositif recevant une représentation tridimensionnelle de l'espace

aérien à un instant où un conflit aérien est détecté entre l'aéronef de référence et l'au moins un autre aéronef, le dispositif étant caractérisé en ce qu'il comprend :

- une unité d'encodage de l'espace aérien configurée pour déterminer une représentation de l'espace aérien en dimension réduite en appliquant un auto-encodeur récurrent à la représentation tridimensionnelle de l'espace aérien à l'instant de détection du conflit aérien
- une unité d'aide à la décision configurée pour déterminer une action de résolution de conflit à mettre en œuvre par l'aéronef de référence, l'unité d'aide à la décision mettant en œuvre un algorithme d'apprentissage par renforcement profond pour déterminer l'action à partir de ladite représentation de l'espace aérien en dimension réduite, d'informations relatives à l'aéronef de référence et/ou à l'au moins un autre aéronef, et d'une géométrie correspondant audit conflit aérien.

[0012] Selon certains modes de réalisation, l'auto-encodeur récurrent peut être préalablement entraîné en utilisant des données réelles des plans de vol de l'aéronef de référence et de l'au moins un autre aéronef.

[0013] Selon certains modes de réalisation, l'auto-encodeur peut être un auto-encodeur LSTM (Long Short-Term Memory).

[0014] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné à approximer, pour une représentation donnée d'un scénario dans l'espace aérien à l'instant où un conflit est détecté, une fonction de récompense, ladite action correspondant à une stratégie optimale maximisant ladite fonction de récompense lors de la phase d'entraînement.

[0015] Selon certains modes de réalisation, la fonction de récompense peut associer une valeur à chaque triplé comprenant une situation aérienne à un premier instant donné, une action prise à un temps donné, et une situation aérienne à un second instant donné, ladite valeur se décomposant en plusieurs pénalités comprenant:

- une pénalité positive si l'action prise à l'instant donné a résolu le conflit, ou
- une pénalité négative si l'action prise à l'instant donné n'a pas permis de résoudre le conflit ou a engendré au moins un autre conflit aérien ;
- une pénalité négative si l'action prise à l'instant donné engendre une nouvelle trajectoire provoquant un détour ;
- une pénalité positive si l'action prise à l'instant donné engendre une nouvelle trajectoire plus courte;
- une pénalité négative si l'action prise à l'instant donné permet de résoudre le conflit aérien et la résolution a lieu proche du conflit ;

- une pénalité négative croissante avec le nombre d'actions prises pour la résolution du conflit aérien.

[0016] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné en utilisant des données opérationnelles et des scénarios correspondant à toutes les manœuvres possibles de l'aéronef de référence, toutes les actions possibles pour résoudre le conflit aérien, et toutes les catégories possibles des aéronefs en conflit.

[0017] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être un réseau de neurones profond mettant en œuvre une technique d'apprentissage par renforcement.

[0018] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être choisi parmi les algorithmes de la famille des algorithmes de Q-learning ou de la famille des algorithmes actor-critic.

[0019] Selon certains modes de réalisation, au moins deux aéronefs parmi l'aéronef de référence et l'au moins un autre aéronef peuvent être de catégories différentes.

[0020] Selon certains modes de réalisation, l'action peut être choisie dans un groupe comprenant la régulation de la vitesse de l'aéronef de référence, le changement d'altitude de l'aéronef de référence, le changement de direction de l'aéronef de référence avec retour sur la trajectoire initiale, le direct sur un point nommé, et l'attente sans faire d'action.

[0021] Les modes de réalisation de l'invention fournissent en outre un procédé pour la gestion de trafic aérien, dans un espace aérien comprenant un aéronef de référence et au moins un autre aéronef, le procédé comprenant une étape pour recevoir une représentation tridimensionnelle de l'espace aérien à un instant où un conflit aérien est détecté entre l'aéronef de référence et l'au moins un autre aéronef, le procédé étant caractérisé en ce qu'il comprend les étapes consistant à :

- déterminer une représentation de l'espace aérien en dimension réduite en appliquant un auto-encodeur récurrent à la représentation tridimensionnelle de l'espace aérien à l'instant de détection de conflit aérien;
- déterminer une action de résolution de conflit à mettre en œuvre par l'aéronef de référence, l'action étant déterminée à partir de la représentation de l'espace aérien en dimension réduite, d'informations relatives à l'aéronef de référence et/ou à l'au moins un autre aéronef, et d'une géométrie correspondant au conflit aérien, en mettant en œuvre un algorithme d'apprentissage par renforcement profond pour déterminer ladite action.

[0022] Avantageusement, les modes de réalisation de l'invention permettent de résoudre des conflits aériens dans un espace aérien tridimensionnel (3D), considérant un nombre d'aéronefs et de couloirs aériens non limités, des géométries de conflits ne se limitant pas aux lignes droites, une hétérogénéité des catégories et de compagnies d'aéronefs, et un nombre élevé d'actions possibles pour la résolution de conflits aériens comprenant la régulation de la vitesse, le changement d'altitude, le changement de direction avec retour sur la trajectoire initiale, la possibilité de couper la route, et la prise d'aucune action (qui est une action en soi). Le choix de l'action effectuée permet de résoudre le conflit aérien tout en tenant compte des autres aéronefs environnants pour éviter de nouveaux conflits et tout en minimisant l'éventuel détour effectué, ce qui permet de réduire la consommation de carburant.

[0023] Avantageusement, les modes de réalisation de l'invention permettent l'aide à la décision pour la résolution de conflits aériens tenant compte des considérations techniques et préférences des contrôleurs aériens et des pilotes pour favoriser certaines actions (par exemple éviter le changement d'altitude dans la mesure du possible).

[0024] Avantageusement, les modes de réalisation de l'invention fournissent une aide à la décision pour la résolution de conflits aériens à moyen terme en utilisant un algorithme d'apprentissage par renforcement profond.

[0025] Avantageusement, l'algorithme d'apprentissage par renforcement selon les modes de réalisation de l'invention se généralise à n'importe quel type de scénario et à des géométries de conflits non précédemment rencontrés sans nécessiter le réentraînement pour chaque type de scénario.

[0026] Avantageusement, l'algorithme d'apprentissage par renforcement selon les modes de réalisation de l'invention met en œuvre un réseau de neurones récurrent pour permettre la résolution de conflits sans limitation du nombre d'aéronefs.

[0027] Avantageusement, l'algorithme d'apprentissage par renforcement selon les modes de réalisation de l'invention tient compte de trois niveaux d'incertitude sur l'impact d'une action possible pour la résolution de conflits aériens.

[0028] Avantageusement, les modes de réalisation de l'invention fournissent une aide à la décision pour les contrôleurs aériens.

Brève description des dessins

[0029] D'autres caractéristiques et avantages de l'invention apparaîtront à l'aide de la description qui suit faite en référence aux dessins annexés, donnés à titre d'exemple, et qui représentent, respectivement :

[0030] - La figure 1 est un schéma représentant un dispositif de gestion de conflit aérien, selon certains modes de réalisation de l'invention.

[0031] - La figure 2 est un logigramme représentant un procédé pour la gestion de conflit aérien, selon certains modes de réalisation de l'invention.

Description détaillée

[0032] Les modes de réalisation de l'invention fournissent un dispositif et un procédé pour la gestion d'un conflit aérien entre un aéronef de référence et au moins un autre aéronef (aussi référencé 'au moins un second aéronef') à partir d'une représentation tridimensionnelle de l'espace aérien à l'instant où le conflit aérien est détecté.

[0033] Les modes de réalisation de l'invention peuvent être utilisés dans les systèmes de contrôles du trafic aérien pour l'aide à la décision aux contrôleurs aériens afin de résoudre les conflits aériens, prévenir les collisions entre aéronefs, et gérer la circulation aérienne.

[0034] Selon les modes de réalisation de l'invention, un aéronef peut être tout type d'aéronef tel qu'un avion, un hélicoptère, une montgolfière, ou un drone.

[0035] Tel qu'utilisé ici, un plan de vol d'un aéronef est une suite de points nommés dans un espace à quatre dimensions comprenant une latitude, une longitude, une altitude, et une valeur temporelle (ou temps de passage ou 'Estimated Time of Overflight' en langue anglo-saxonne). Les points nommés représentent la trajectoire que doit suivre l'aéronef aux temps indiqués par les valeurs temporelles.

[0036] Tel qu'utilisé ici, un scénario représente un ensemble de plans de vols avec les identifiants et les catégories d'au moins un aéronef.

[0037] Selon certains modes de réalisation, deux aéronefs parmi l'aéronef de référence et l'au moins un autre aéronef peuvent être de catégories différentes.

[0038] Selon certains modes de réalisation dans lesquels l'aéronef de référence et l'au moins un autre aéronef sont des avions, l'aéronef de référence et au moins un autre aéronef peuvent être de compagnies d'avions différentes.

[0039] Selon certains modes de réalisation, l'aéronef de référence peut être préalablement sélectionné de façon aléatoire.

[0040] En référence à la figure 1, les modes de réalisation de l'invention fournissent un dispositif 100 pour la gestion d'un conflit aérien entre un aéronef de référence et au moins un autre aéronef à partir d'une représentation tridimensionnelle de l'espace aérien à l'instant où le conflit aérien est détecté.

[0041] Selon certains modes de réalisation, le dispositif 100 peut comprendre une unité d'encodage de l'espace aérien 101 configurée pour déterminer une représentation de l'espace aérien en dimension réduite en appliquant un auto-encodeur récurrent à la représentation tridimensionnelle de l'espace aérien à l'instant de détection du conflit aérien, l'encodage de l'espace aérien correspondant à l'aéronef de référence et à l'au moins un autre aéronef impliqués dans le conflit aérien. L'auto-encodeur récurrent est un réseau de neurones artificiels utilisé pour apprendre une représentation (ou encodage) d'un ensemble de données dans le but de réduire la dimension de cet ensemble.

[0042] Selon certains modes de réalisation, l'auto-encodeur récurrent peut être préalablement entraîné en utilisant des données réelles des plans de vol de l'aéronef de référence et de l'au moins un autre aéronef, indépendamment de la résolution du conflit aérien. Cette phase d'entraînement peut être effectuée hors ligne en utilisant une variante de la rétro-propagation comme la méthode du gradient conjugué ou l'algorithme du gradient. L'aspect récurrent de l'auto-encodeur permet avantageusement de gérer un nombre variable d'aéronefs et de ne pas faire dépendre l'architecture du réseau de neurones du nombre d'aéronefs présents simultanément dans l'espace aérien.

[0043] Selon certains modes de réalisation, l'auto-encodeur peut être un auto-encodeur LSTM (acronyme de 'Long Short-Term Memory' en langage anglo-saxon).

[0044] Selon certains modes de réalisation, le dispositif 100 peut en outre comprendre une unité d'aide à la décision 103 configurée pour fournir une action à mettre en œuvre par l'aéronef de référence pour résoudre le conflit aérien, l'unité d'aide à la décision 103 appliquant un algorithme d'apprentissage par renforcement profond pour déterminer l'action à partir de la représentation de l'espace aérien en dimension réduite fournie par l'auto-encodeur, d'informations relatives à l'aéronef de référence et/ou à l'au moins un autre aéronef, et de la géométrie correspondant au conflit aérien.

[0045] Selon certains modes de réalisation, les informations relatives à l'aéronef de référence et/ou à l'au moins un autre aéronef peuvent comprendre la distance verticale, la distance horizontale, et l'azimut entre l'aéronef de référence et l'au moins un autre aéronef. Les informations peuvent en outre comprendre les distances et les angles entre l'aéronef de référence et au moins un aéronef non inclus dans le conflit aérien, ainsi que la catégorie de l'aéronef de référence et la position des derniers points nommés.

[0046] Selon certains modes de réalisation, l'action peut être choisie dans un groupe comprenant la régulation de la vitesse de l'aéronef de référence, le changement d'altitude de l'aéronef de référence, le changement de direction de l'aéronef de référence avec retour sur la trajectoire initiale, le direct sur un point nommé, l'attente sans faire d'action.

[0047] Selon les modes de réalisation de l'invention, l'unité d'aide à la décision 103 se base sur des techniques d'apprentissage par renforcement profond combinant l'apprentissage par renforcement avec les réseaux de neurones artificiels pour déterminer, à partir de l'encodage de l'espace aérien au moment du conflit aérien, l'action optimale à mettre en œuvre par l'aéronef de référence pour résoudre le conflit aérien tout en tenant compte d'un ensemble de contraintes. L'ensemble de contrainte selon les modes de réalisation de l'invention comprend :

- la gestion de l'espace aérien tridimensionnel ;
- la gestion de tous les types d'actions possibles pour la résolution de conflits aériens ;
- la gestion d'un nombre variable d'aéronefs avec hétérogénéité des catégories et des compagnies ;
- la résolution du conflit aérien avec prise en compte des aéronefs alentour pour éviter la création de nouveaux conflits aériens, et
- la résolution efficace du conflit aérien tout en minimisant le détour effectué suite à une prise d'action, et la prise en compte de scénarios et de géométries de conflits non précédemment rencontrés.

[0048] L'apprentissage par renforcement consiste, pour un agent autonome, à apprendre les actions à prendre, à partir d'expériences, de façon à optimiser une fonction de récompense quantitative au cours du temps.

[0049] L'agent autonome est plongé dans un environnement et prend ses décisions en fonction de son état courant. En retour, l'environnement procure à l'agent autonome une récompense, qui est une valeur numérique qui peut être positive, négative, ou nulle, les récompenses positives mettant l'accent sur une action désirée, les récompenses négatives mettant l'accent sur une action dont l'agent doit s'éloigner, et les récompenses nulles indiquant que l'action est neutre. L'environnement peut changer au fur et à mesure que l'agent prend des actions, les actions étant les méthodes de l'agent lui permettant d'interagir et de changer son environnement, et donc de se transférer entre états.

[0050] L'agent autonome, cherche, au travers d'expériences itérées, un comportement décisionnel (aussi appelé 'stratégie' ou 'politique') optimal permettant la maximisation des récompenses au cours du temps.

[0051] La base du modèle d'apprentissage par renforcement consiste ainsi en :

- un ensemble d'états S de l'agent dans l'environnement ;
- un ensemble d'actions A que l'agent peut effectuer, et
- un ensemble de valeurs scalaires R (aussi appelées récompenses ou fonction de récompenses) que l'agent peut obtenir de la part de l'environnement. Chaque fonction de récompense reflète le comportement que doit adopter l'agent.

[0052] A chaque pas de temps t de l'algorithme d'apprentissage par renforcement, l'agent perçoit son état $s_t \in S$ (aussi appelée situation à l'instant donné t) et l'ensemble des actions possibles $A(s_t)$. L'agent choisit une action $a \in A(s_t)$ (aussi appelée l'action prise à l'instant donné t) et reçoit de l'environnement un nouvel état s_{t+1} (aussi appelé la situation à l'instant donné $t+1$) et une récompense R_{t+1} . La décision de l'action à choisir par l'agent est prise par une politique $\pi: S \rightarrow A$ qui est une fonction qui conditionnellement à un état, associe une probabilité de sélection à chaque action dans cet état. Le but de l'agent est de maximiser les récompenses globales qu'il reçoit de la part de l'environnement au cours d'un épisode, un épisode comprenant tous les états de l'agent qui se situent entre un état initial et un état terminal. La valeur désignée par Q-valeur et notée $Q(s, a)$, mesure la récompense globale attendue si l'agent est dans l'état $s \in S$ et effectue l'action a , puis continue à interagir avec son environnement jusqu'à la fin de l'épisode courant selon une politique π .

[0053] Selon les modes de réalisation de l'invention :

- chaque aéronef est un agent autonome qui doit apprendre à résoudre les conflits dans l'espace aérien ;
- l'environnement de l'agent est une représentation de l'espace aérien décrite par un scénario, et
- les actions prises par un aéronef comprennent toutes les actions possibles de contrôle aérien comprenant le changement de direction, le changement d'altitude, le changement de vitesse, le direct sur un point nommé, et le changement de direction avec retour sur la trajectoire initiale.

[0054] Selon certains modes de réalisation, l'agent peut ne pas observer la totalité de l'environnement mais seulement quelques variables lui permettant d'évoluer efficacement dans l'environnement. Ces variables peuvent comprendre la vitesse, la position, et l'altitude de l'agent et de tous les autres aéronefs présents, ainsi que des informations sur le conflit aérien à résoudre et les positions des points nommés sur lesquels l'agent peut faire un 'direct'.

[0055] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné à approximer, pour une représentation donnée du scénario dans l'espace aérien à l'instant de détection d'un conflit, une fonction de récompense, de manière à ce que l'action (optimale) à mettre en œuvre par l'aéronef de référence corresponde à la stratégie optimale apprise permettant la maximisation de la fonction de récompense. L'entraînement de l'algorithme d'apprentissage par renforcement permet ainsi de déterminer les sommes cumulées (ou des récompenses globales) futures que peut obtenir l'agent pour une action et une situation (ou scénario) donnés. Après

l'entraînement et la convergence de l'algorithme d'apprentissage par renforcement, l'action qui rapporte la fonction de récompense maximale peut être fournie à l'aéronef de référence afin de suivre la stratégie optimale pour la résolution du conflit aérien.

[0056] Selon certains modes de réalisation, la fonction de récompense peut être préalablement modélisée afin que la stratégie optimale de maximisation des récompenses corresponde à l'ensemble des contraintes précédemment défini. Selon certains modes de réalisation, la fonction de récompense peut être modélisée pour associer une valeur à chaque triplé comprenant une situation aérienne à un premier instant donné t , une action a prise à un temps donné t , et une situation aérienne à un second instant donné $t+1$, la valeur reflétant l'attractivité du triplé et se décomposant en plusieurs pénalités comprenant :

- une pénalité positive si l'action a prise à l'instant donné t a résolu le conflit
- une pénalité négative si l'action a prise à l'instant donné t n'a pas permis de résoudre le conflit ou a engendré au moins un autre conflit aérien
- une pénalité négative si l'action a prise à l'instant donné t engendre une nouvelle trajectoire provoquant un détour
- une pénalité positive si l'action a prise à l'instant donné t engendre une nouvelle trajectoire plus courte
- une pénalité négative si l'action a prise à l'instant donné t permet de résoudre le conflit aérien et la résolution a lieu proche du conflit, et
- une pénalité négative croissante avec le nombre d'actions prises pour la résolution du conflit aérien.

[0057] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné en utilisant des données opérationnelles et des scénarios correspondant à toutes les manœuvres possibles de l'aéronef de référence, toutes les actions possibles pour résoudre un conflit aérien, et toutes les catégories possibles des aéronefs en conflit.

[0058] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné en utilisant des scénarios réalistes créés automatiquement à partir de données opérationnelles et en réalisant l'augmentation des données pour un apprentissage profond par exemple en variant les catégories des aéronefs, en ajoutant des délais sur certains aéronefs permettant de modifier et rajouter des conflits aériens.

[0059] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné en utilisant des données générées par des

dispositifs de détection de conflits et/ou des dispositifs de modification de trajectoires (non illustrés sur la figure 1).

[0060] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être un réseau de neurones profond mettant en œuvre une technique d'apprentissage par renforcement.

[0061] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être choisi parmi les algorithmes de la famille des algorithmes de Q-learning ou de la famille des algorithmes actor-critic.

[0062] En référence à la figure 2, les modes de réalisation de l'invention fournissent en outre un procédé pour la gestion d'un conflit aérien entre un aéronef de référence et au moins un autre aéronef à partir d'une représentation tridimensionnelle de l'espace aérien à l'instant où le conflit aérien est détecté.

[0063] A l'étape 201, une représentation tridimensionnelle de l'espace aérien au moment du conflit aérien peut être reçue.

[0064] A l'étape 203, une représentation de l'espace aérien en dimension réduite peut être déterminée en appliquant un auto-encodeur récurrent à la représentation tridimensionnelle de l'espace aérien à l'instant de détection du conflit aérien, l'encodage de l'espace aérien correspondant à l'aéronef de référence et à l'au moins un autre aéronef impliqués dans le conflit aérien.

[0065] Selon certains modes de réalisation, l'étape 203 peut comprendre une sous-étape effectuée hors ligne pour entraîner l'auto-encodeur récurrent en utilisant des données réelles des plans de vol de l'aéronef de référence et de l'au moins un autre aéronef, indépendamment de la résolution du conflit aérien.

[0066] Selon certains modes de réalisation, l'auto-encodeur récurrent peut être entraîné en utilisant une variante de la rétro-propagation comme la méthode du gradient conjugué ou l'algorithme du gradient.

[0067] Selon certains modes de réalisation, l'auto-encodeur récurrent peut être un auto-encodeur LSTM.

[0068] A l'étape 205, une action à mettre en œuvre par l'aéronef de référence peut être déterminée à partir de la représentation de l'espace aérien en dimension réduite, d'informations relatives à l'aéronef de référence et/ou l'au moins un autre aéronef, et de la géométrie du conflit aérien, en appliquant un algorithme d'apprentissage par renforcement profond.

[0069] Selon certains modes de réalisation, les informations relatives à l'aéronef de référence et/ou l'au moins un autre aéronef peuvent comprendre la distance verticale, la distance horizontale, et l'azimut entre l'aéronef de référence et l'au moins un autre aéronef. Les informations peuvent en outre comprendre les distances et les angles entre l'aéronef de référence et au moins un aéronef non inclus dans le conflit aérien, ainsi que la catégorie de l'aéronef de référence et la position des derniers points nommés.

[0070] Selon certains modes de réalisation, l'action à mettre en œuvre par l'aéronef de référence peut être choisie dans un groupe comprenant la régulation de la vitesse de l'aéronef de référence, le changement d'altitude de l'aéronef de référence, le changement de direction de l'aéronef de référence avec retour sur la trajectoire initiale, le direct sur un point nommé, l'attente sans faire d'action.

[0071] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être conçu pour déterminer l'action optimale parmi toutes les actions possibles pour la résolution de conflits aériens en respectant un ensemble de contraintes ou exigences comprenant :

- la gestion de l'espace aérien tridimensionnel ;
- la gestion de tous les types d'actions possibles pour la résolution de conflits aériens ;
- la gestion d'un nombre variable d'aéronefs avec hétérogénéité des catégories et des compagnies ;
- la résolution du conflit aérien avec prise en compte des aéronefs alentour pour éviter la création de nouveaux conflits aériens, et
- la résolution efficace du conflit aérien tout en minimisant le détour effectué suite à une prise d'action, et la prise en compte de scénarios et de géométries de conflits non précédemment rencontrés.

[0072] Selon les modes de réalisation de l'invention, le modèle de l'algorithme d'apprentissage par renforcement profond peut être définie par :

- un agent autonome correspondant à un aéronef, l'agent autonome devant apprendre les actions à prendre pour résoudre les conflits dans l'espace aérien à partir d'expérience de façon à optimiser une fonction de récompense au cours du temps ;
- l'environnement de l'agent correspond à une représentation de l'espace aérien décrite par un scénario, l'agent étant plongé dans cet environnement et prenant des actions lui permettant d'interagir et de changer son environnement et de changer d'états ;
- les actions prises par un agent comprennent toutes les actions possibles de contrôle aérien que peut prendre un aéronef pour résoudre un conflit aérien, comprenant le changement de

direction, le changement d'altitude, le changement de vitesse, le direct sur un point nommé, et le changement de direction avec retour sur la trajectoire initiale.

[0073] Selon certains modes de réalisation, l'agent peut ne pas observer la totalité de l'environnement mais seulement quelques variables lui permettant d'évoluer efficacement dans l'environnement. Ces variables peuvent comprendre la vitesse, la position, et l'altitude de l'agent et de tous les autres aéronefs présents, ainsi que des informations que le conflit aérien à résoudre et les positions des points nommés sur lesquels l'agent peut faire un 'direct'.

[0074] A chaque pas de temps t de l'algorithme d'apprentissage par renforcement, l'agent perçoit son état $s_t \in S$ à l'instant donné t et l'ensemble des actions possibles $A(s_t)$. L'agent choisit une action $a \in A(s_t)$ et reçoit de l'environnement un nouvel état s_{t+1} correspondant à la situation à l'instant donné $t+1$ et une récompense R_{t+1} . La décision de l'action à choisir par l'agent est prise par une politique $\pi: S \rightarrow A$ qui est une fonction qui conditionnellement à un état, associe une probabilité de sélection à chaque action dans cet état. Le but de l'agent est de maximiser les récompenses globales qu'il reçoit de la part de l'environnement au cours d'un épisode, un épisode comprenant tous les états de l'agent qui se situent entre un état initial et un état terminal. La valeur désignée par Q-valeur et notée $Q(s, a)$, mesure la récompense globale attendue si l'agent est dans l'état $s \in S$ et effectue l'action a , puis continue à interagir avec son environnement jusqu'à la fin de l'épisode courant selon une politique π .

[0075] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné à approximer, pour une représentation donnée du scénario dans l'espace aérien au moment d'un conflit, une fonction de récompense, de manière à ce que l'action à mettre en œuvre par l'aéronef de référence corresponde à la stratégie optimale apprise permettant la maximisation de la fonction de récompense. L'entraînement de l'algorithme d'apprentissage par renforcement permet ainsi de déterminer les sommes cumulées (ou des récompenses globales) futures que peut obtenir l'agent pour une action et une situation (ou scénario) donnés. Après l'entraînement et la convergence de l'algorithme d'apprentissage par renforcement, l'action qui rapporte la fonction de récompense maximale pour la situation courante donnée au moment du conflit peut être sélectionnée, elle représente l'action qui correspond à la stratégie optimale pour la résolution du conflit aérien.

[0076] Selon certains modes de réalisation, la fonction de récompense peut être préalablement modélisée afin que la stratégie optimale de maximisation des récompenses corresponde à l'ensemble des contraintes précédemment défini. Selon certains modes de réalisation, la fonction de récompense peut être modélisée pour associer une valeur à chaque

triplé comprenant une situation aérienne à un premier instant donné t , une action a prise à un temps donné t , et une situation aérienne à un second instant donné $t+1$, la valeur reflétant l'attractivité du triplé et se décomposant en plusieurs pénalités comprenant :

- une pénalité positive si l'action a prise à l'instant donné t a résolu le conflit ;
- une pénalité négative si l'action a prise à l'instant donné t n'a pas permis de résoudre le conflit ou a engendré au moins un autre conflit aérien ;
- une pénalité négative si l'action a prise à l'instant donné t engendre une nouvelle trajectoire provoquant un détour ;
- une pénalité positive si l'action a prise à l'instant donné t engendre une nouvelle trajectoire plus courte ;
- une pénalité négative si l'action a prise à l'instant donné t permet de résoudre le conflit aérien et la résolution a lieu proche du conflit, et
- une pénalité négative croissante avec le nombre d'actions prises pour la résolution du conflit aérien.

[0077] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné en utilisant des données opérationnelles et des scénarios correspondant à toutes les manœuvres possibles de l'aéronef de référence, toutes les actions possibles pour résoudre un conflit aérien, et toutes les catégories possibles des aéronefs en conflit.

[0078] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être préalablement entraîné en utilisant des scénarios réalistes créés automatiquement à partir de données opérationnelles et en réalisant l'augmentation des données pour un apprentissage profond par exemple en variant les catégories des aéronefs, en ajoutant des délais sur certains aéronefs permettant de modifier et rajouter des conflits aériens.

[0079] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être un réseau de neurones profond mettant en œuvre une technique d'apprentissage par renforcement.

[0080] Selon certains modes de réalisation, l'algorithme d'apprentissage par renforcement profond peut être choisi parmi les algorithmes de la famille des algorithmes de Q-learning ou de la famille des algorithmes actor-critic.

[0081] L'invention fournit en outre un produit programme d'ordinateur pour la gestion d'un conflit aérien entre un aéronef de référence et au moins un autre aéronef à partir d'une représentation tridimensionnelle de l'espace aérien à l'instant où le conflit aérien est détecté, le produit programme d'ordinateur comprenant des instructions de code de programme

informatique qui, lorsqu'elles sont exécutées par un ou plusieurs processeurs, amènent le ou les processeurs à :

- déterminer une représentation de l'espace aérien en dimension réduite en appliquant un auto-encodeur récurrent à la représentation tridimensionnelle de l'espace aérien à l'instant de détection du conflit aérien;
- déterminer une action à mettre en œuvre par l'aéronef de référence à partir de la représentation de l'espace aérien en dimension réduite, d'informations relatives à l'aéronef de référence et/ ou l'au moins un autre aéronef, et de la géométrie du conflit aérien, en appliquant un algorithme d'apprentissage par renforcement profond.

[0082] En général les routines exécutées pour mettre en œuvre les modes de réalisation de l'invention, qu'elles soient mises en œuvre dans le cadre d'un système d'exploitation ou d'une application spécifique, d'un composant, d'un programme, d'un objet, d'un module ou d'une séquence d'instructions, ou même d'un sous-ensemble de ceux-là, peuvent être désignées comme "code de programme informatique" ou simplement "code de programme". Le code de programme comprend typiquement des instructions lisibles par ordinateur qui résident à divers moments dans des dispositifs divers de mémoire et de stockage dans un ordinateur et qui, lorsqu'elles sont lues et exécutées par un ou plusieurs processeurs dans un ordinateur, amènent l'ordinateur à effectuer les opérations nécessaires pour exécuter les opérations et/ou les éléments propres aux aspects variés des modes de réalisation de l'invention. Les instructions d'un programme, lisibles par ordinateur, pour réaliser les opérations des modes de réalisation de l'invention peuvent être, par exemple, le langage d'assemblage, ou encore un code source ou un code objet écrit en combinaison avec un ou plusieurs langages de programmation.

[0083] L'invention n'est pas limitée aux modes de réalisation décrits ci-avant à titre d'exemple non limitatif. Elle englobe toutes les variantes de réalisation qui pourront être envisagées par l'homme du métier.