Validation du regroupement d'occurrences vidéo de personnes

5.1 Présentation des expérimentations

Après avoir présenté de façon théorique nos propositions pour le regroupement de personnes, nous allons les valider de façon expérimentale. Pour cela, nous allons utiliser le corpus de vidéos issu du défi ANR REPERE (présenté dans le contexte de la thèse), proposant de plus de 100 vidéos d'émissions audiovisuelles, dans lesquelles les personnes ont été annotées de façon manuelle.

Dans un premier temps, nous allons vérifier que les résultats de mise en correspondance entre histogrammes spatio-temporels que nous obtenons ont du sens. Un test statistique confirme qu'il y a une différence significative dans la similarité entre des histogrammes spatio-temporels d'occurrences vidéo de même personne et d'histogrammes spatio-temporels de personnes différentes. Cela montre que les résultats ne sont pas obtenus de façon aléatoire et que notre approche permet effectivement de discriminer les occurrences vidéo de personnes.

Nous allons ensuite regarder l'évolution de la précision de notre système en fonction du paramétrage en cherchant à identifier l'espace de couleur le plus approprié, le nombre de partitions optimal, ainsi que la stratégie de construction la plus adaptée.

Une fois ces paramètres déterminés, nous identifions ces mêmes paramètres pour différentes approches de l'état de l'art comme les histogrammes de couleurs, les spatiogrammes et les histogrammes de LBP. Nous comparerons les résultats obtenus, pour une tâche de recherche, dans les différents cas avec ceux obtenus avec notre approche.

Les mesures de similarités données par les différentes approches seront ensuite utilisées pour effectuer le regroupement d'occurrences vidéo de personnes. Les différents groupes obtenus seront évalués selon de nombreux critères afin de déterminer quelle approche convient le mieux pour selon l'application considérée.

5.2 Présentation des données de test

Le corpus de données fourni pour le défi ANR REPERE consiste en plusieurs heures d'émissions télévisées annotées partiellement. Ces données viennent de deux chaînes télévisées françaises : LCP et BFMTV. Plusieurs émissions de ces chaînes sont présentes

dans le corpus, elles ont des longueurs variables et la façon dont chaque émission est filmée varie aussi. Certaines contiennent des plans filmés en extérieur.

Les données sont encodées au format vidéo MPEG avec une taille, à l'affichage (*Display Aspect Ratio*), de 720x576 pixels. En revanche, dans le cas de la chaîne LCP, les vidéos sont encodées avec une taille de 544x576 pixels (*Storage Aspect Ratio*) qui doit être redimensionnée en 720x576 pixels pour obtenir le ratio original de l'image.

Les annotations sont fournies dans des fichiers XML en utilisant le schéma de données du logiciel VIPER (*VIdeo Performance Evaluation Resource*) [1]. Les annotations ne concernent pas les vidéos entières, mais uniquement un certain nombre de segments. Un segment annoté pour une personne débute sur l'apparition à l'image d'une personne et termine lors de sa disparition. Pour chaque de segment, une trame clef a été sélectionnée par l'annotateur. Cette trame est choisie aléatoirement avec pour contrainte d'éviter les trames situées à la limite de deux plans. Si cette trame clef contient le visage d'une personne annotée pour ce segment, il est détouré par un polygone, dessiné manuellement par l'annotateur. La quantité d'annotation de personnes dans chaque vidéo varie entre 30% et 90% de la longueur totale de l'émission.

En utilisant les vidéos d'origine et les annotations, nous avons extrait des occurrences vidéo de personnes dont l'identité est connue. En effet, toutes les personnes des vidéos du corpus ne sont pas annotées, c'est le cas notamment des personnes au sein d'une foule ou du public. La plupart des personnes sont présentes dans de nombreuses occurrences vidéo réparties le long de la vidéo. Ceci permet d'établir une collections de tests conséquente qui nous servira de vérité terrain lors de nos expérimentations.

Au total, le corpus est composé de 303 personnes différentes, dont l'identité est donnée par les annotations. Chaque personne apparaît en moyenne dans 15 émissions différentes. Les présentateurs apparaissent naturellement plus fréquemment que les autres personnes : ils peuvent apparaître dans plus de 50 occurrences vidéo par émission alors que certaines personnes peuvent n'apparaître qu'une seule fois.

5.3 Prétraitements des données

Les occurrences vidéo de personnes sont extraites de 141 émissions différentes. Les annotations ont été utilisées pour vérifier que chaque occurrence vidéo de personne contienne au plus une personne. Soulignons que tous les visages présents dans un segment annoté ne sont pas annotés dans le corpus REPERE, selon des critères de tailles et de sémantique. C'est le cas des scènes avec un public, notamment dans les scènes en extérieur. Nous avons filtré manuellement les occurrences vidéo pour nous assurer de la qualité du corpus. Ceci nous permet d'éviter toute confusion entre ces personnes lors de l'évaluation. Car bien que les visages soient annotés en position sur les trames clefs, les segments annotés ne tiennent pas compte du changement de plan. Il n'y a donc aucune garantie de la correspondance des visages en dehors des trames clefs. De plus, tous les visages ne sont pas annotés, même sur les trames clefs.

Ensuite, un algorithme combinant de la détection de visages et les annotations a été utilisé pour retirer toutes les occurrences qui pourraient contenir des personnes non-annotées. Le détecteur nous permet de mettre en évidence toutes les séquences vidéo dont deux visages ou plus ont été détectés dans pour une même trame.

1. Le logiciel VIPER est disponible à l'url http ://viper-toolkit.sourceforge.net/.

Ainsi, à la fin du processus de sélection, nous obtenons 5279 occurrences vidéo de 303 personnes différentes. Chacune est présente en moyenne dans 5 émissions. Les journalistes sont plus représentés que les invités.

5.4 Calcul des matrices de similarités

Chaque occurrence vidéo de personne de notre corpus a été utilisée pour construire des histogrammes spatio-temporels, spatiogrammes et histogrammes de couleur afin de comparer ces trois descripteurs. Ces descripteurs ont été construits en utilisant différentes combinaisons de paramètres :

— nombre de partitions différents (10, 50, 100, 150, 200, 250, 300, 350, 400, 500, 800, 1.000, 1.500, 2.000, 2.500, 5.000, 10.000 et 100.000),
— des espaces de représentation des couleurs différents (RGB, OHTA, HSV),
— des stratégies de constructions différentes (accumulation, fenêtres glissantes, fenêtres sautantes et séparation des canaux).

Les matrices de similarités ont été générées en utilisant des mesures de similarités différentes (χ^2, Bhattacharyya, Bhattacharyya combinée à Mahalanobis et χ^2 combiné à Mahalanobis).

5.5 Paramétrage des HST

Afin de paramétrer de façon optimale les histogrammes spatio-temporels, nous avons exploré de nombreuses combinaison de paramètres. Pour les histogrammes spatio-temporels deux paramètres entre en jeu lors de la construction : le nombre de partitions de l'histogramme et l'espace de représentation des couleurs. Ces deux paramètres auront un impact sur la qualité du regroupement final. Il est important de noter que le coût calculatoire de la mesure de similarité dépend du nombre de partitions.

5.5.1 Variation du nombre de partitions

Dans cette expérimentation, nous faisons varier graduellement le nombre de partitions B des histogrammes spatio-temporels, construits sur l'espace de couleur RGB et nous observons l'impact sur la précision mesurée (cf. Équation 4.20 de la Section 4.7).

Pour mémoire, notre mesure de similarité entre histogrammes spatio-temporels consiste en une similarité de Mahalanobis pondérée par la similarité du χ^2 (cf. Équation 4.7).

Nous avons testé différentes valeurs prises dans l'intervalle $[10; 100.000]$. Cela nous permet d'avoir une bonne résolution du comportement de la précision lors de ses plus grandes variations. L'augmentation du nombre de partitions a été arrêtée après que la précision mesurée a commencé à diminuer. Nous avons choisi de prendre la dernière mesure loin de ce point d'inflexion afin de confirmer le comportement de la courbe.

Dans les résultats de notre expérimentation, présentés dans la Figure 5.1, nous remarquons que la précision augmente rapidement jusqu'à environ 500 partitions. Au-delà de ce seuil, la précision continue d'augmenter mais de moins en moins rapidement, jusqu'à atteindre un point d'arrêt autour de 2.000 partitions pour entamer une diminution de la précision. soulignons que la courbe ne donne la précision que dans l'intervalle des partitions $[10; 10.000]$ [2], nous avons calculé la précision jusqu'à 100.000 partitions. Cette

2. La courbe devenait peu lisible en allant jusqu'à 100.000 partitions.

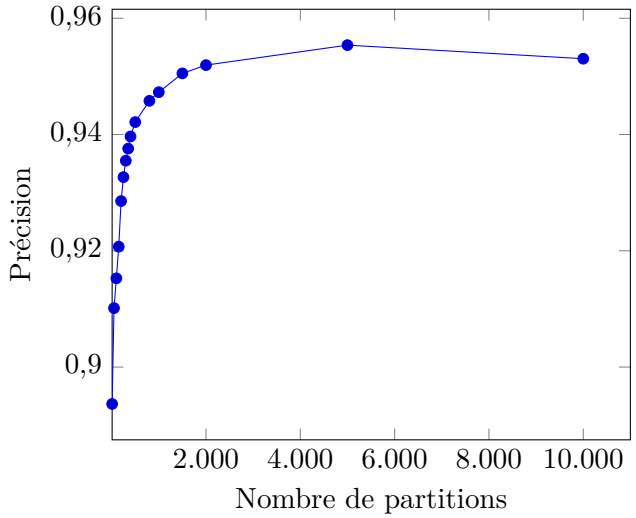

FIGURE 5.1 – Évolution de la précision en fonction du nombre de partitions des HST entre 10 et 10.000 partitions.

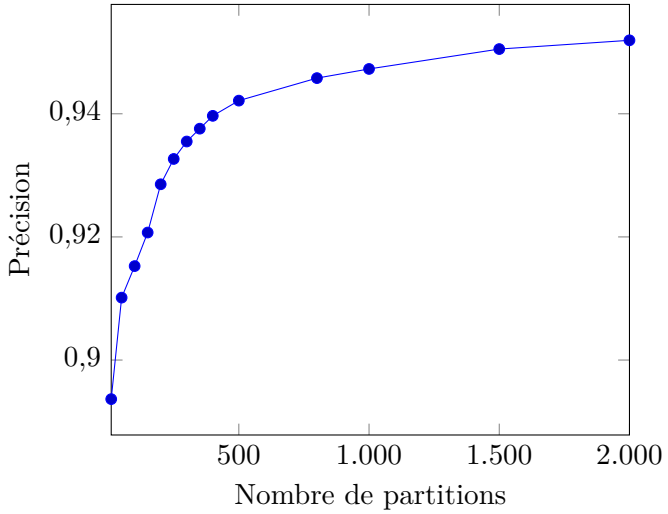

FIGURE 5.2 – Évolution de la précision en fonction du nombre de partitions des HST entre 10 et 2.000 partitions.

diminution se confirme effectivement au-delà de 10.000 partitions.

Ce seuil est particulièrement visible dans la Figure 5.2, qui montre l'évolution de la précision entre 10 et 2.000 partitions.

Une différence de 6 points de base (de 0,89 à 0,95) sur la précision peut sembler faible. Quand on rapporte cela à notre application de recherche d'occurrences vidéo de personnes, ce sont presque 320 occurrences supplémentaires qui sont correctement renvoyées par le système. Cette différence n'est pas anodine et mérite l'effort supplémentaire, en complexité, à fournir.

Ainsi, l'augmentation du nombre de partitions permet de mieux discriminer les oc-

currences vidéo de personnes jusqu'à une certaine limite. Une fois cette limite atteinte, on suppose que l'information est diluée dans plusieurs partitions et perd en cohérence. Dès lors, augmenter le nombre de partitions ne contribue qu'à faire diminuer la précision tout en augmentant le coût calculatoire des histogrammes spatio-temporels.

5.5.2 Variation de l'espace de couleurs

Nous allons maintenant étudier l'évolution de cette même précision quand l'espace de représentation des couleurs varie (le nombre de partitions restant constant). Cela va nous permettre d'étudier le comportement des histogrammes spatio-temporels sur différents espaces de couleurs. Les espaces de couleurs HSV, OHTA et RGB sont comparés.

Le nombre de partitions a été fixé à 350. Ce nombre de partitions, permet de comparer la précision sur les différents espaces de représentation avant le seuil de 2.000 partitions, au-delà duquel la précision commence à diminuer. En nous plaçant suffisamment loin de ce seuil, nous savons que la précision est plus volatile, l'impact du choix de l'espace de représentation sera ainsi mieux mis en évidence.

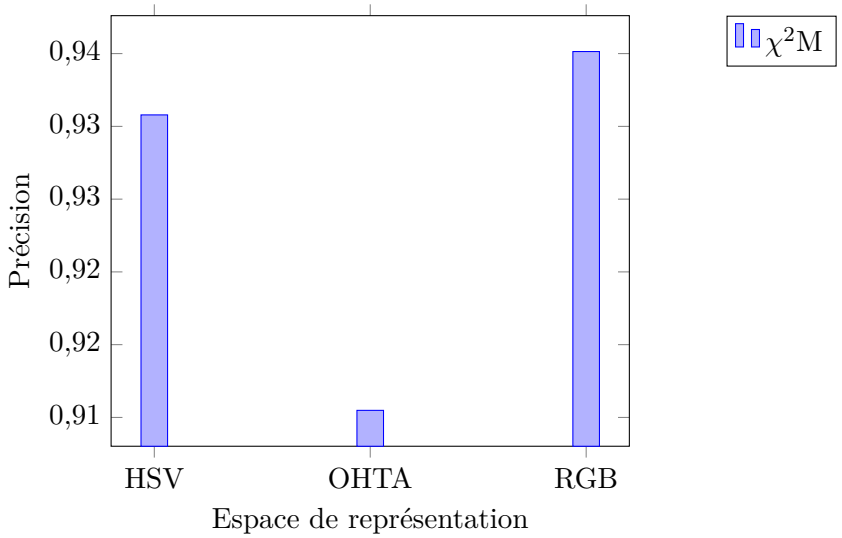

FIGURE 5.3 – Précision pour des espaces de représentation des couleurs différents et un nombre de partitions fixé à 350.

Dans la Figure 5.3, on remarque que l'espace de représentation OHTA donne les moins bons résultats. Les espaces de couleur HSV et RGB donnent des précisions proches, bien supérieures à celles obtenues OHTA. RGB obtient la meilleure précision.

Le descripteur de couleur OHTA produit des résultats inférieurs à ceux obtenus sur les espaces HSV ou RGB. Cela est probablement dû à la construction même de cet espace de couleur, ayant pour objectif de réduire au maximum la corrélation entre les différents canaux. Les canaux n'étant pas corrélés, le découpage linéaire de l'espace de couleur formé par les trois canaux est moins pertinent. Il serait plus intéressant d'exploiter les différents canaux de l'espace de couleur OHTA et de les décrire séparément. Néanmoins, cela aurait pour conséquence de tripler le coût de la construction et de la comparaison.

De façon générale, le descripteur couleur RGB donne les meilleurs résultats en termes de précision et ne présente aucun surcoût, les images étant par convention matérielle

exprimées dans cet espace de couleur. Dans le cas d'autres espaces, une conversion depuis RGB est nécessaire. Comme nous l'avons présenté dans l'état de l'art sur les espaces de couleurs (Section 2.2.4), cette conversion peut être très coûteuse à calculer car elle dépend du nombre de pixels ainsi que du nombre de trames sur lesquelles sont construits les histogrammes spatio-temporels.

5.5.3 Comparaison avec un descripteur de textures

La comparaison entre des histogrammes spatio-temporels construits sur la représentation de couleur RGB et la représentation de texture LBP a été réalisée. Comme nous l'avons présenté dans l'état de l'art sur la ré-identification de personnes, plusieurs approches exploitent des histogrammes de LBP [53, 97, 15, 84].

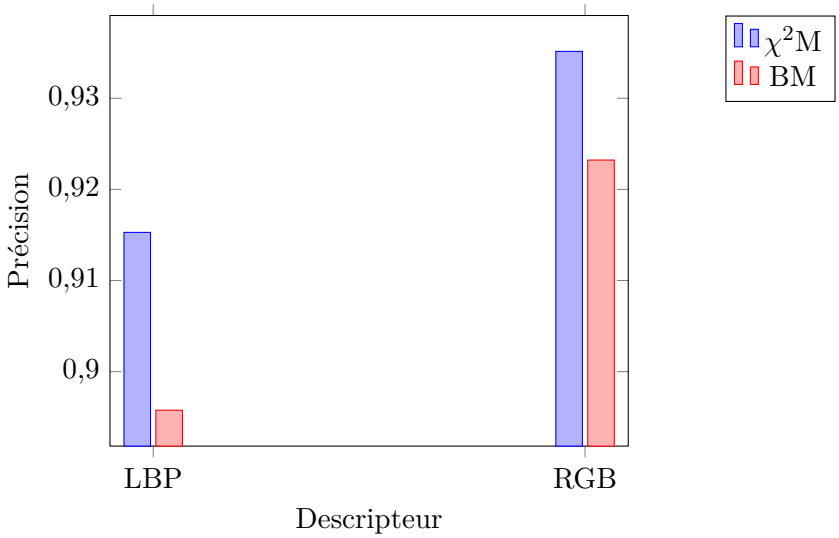

FIGURE 5.4 – Précision entre des histogrammes spatio-temporels construits sur l'espace de couleur RGB et le descripteur de texture LBP pour un nombre de partitions constant, fixé à 350, selon la mesure de similarité considéré.

La précision obtenue, présentée dans la Figure 5.4, avec l'approche exploitant la représentation de la texture par le descripteur LBP offre des résultats inférieurs à ceux obtenus par les espaces de représentation des couleurs RGB. Cela peut être dû aux informations de textures qui à l'échelle du corps complet de la personne sont moins pertinentes pour distinguer les personnes.

De plus, le descripteur LBP nécessite une étape d'extraction préalable, sur chaque trame, avant qu'un histogramme spatio-temporel puisse être construit dessus. Cette étape de calcul introduit un coût calculatoire important. Ainsi les histogrammes spatio-temporels basés sur le descripteur LBP sont non seulement moins précis pour mettre en correspondance des occurrences vidéo de personnes mais leur coût calculatoire est supérieur aux approches basées sur les espaces de couleurs.

5.5.4 Comparaison des mesures de similarités

Après avoir étudié l'évolution de la précision en fonction du nombre de partitions et de l'espace de représentation des couleurs. Nous allons maintenant comparer la précision mesurée en fonction de la mesure de similarité exploitée. Ainsi, nous allons nous intéresser aux résultats des combinaisons Bhattacharyya-Mahalanobis (BM) et χ^2-Mahalanobis (χ^2M). En premier lieu, nous rappelons que la complexité des deux mesures est équivalente, comme nous l'avons mentionné dans la Section 4.5. Dans un premier temps nous allons faire varier l'espace de couleur utilisé, en gardant le nombre de partitions constant, fixé à 350.

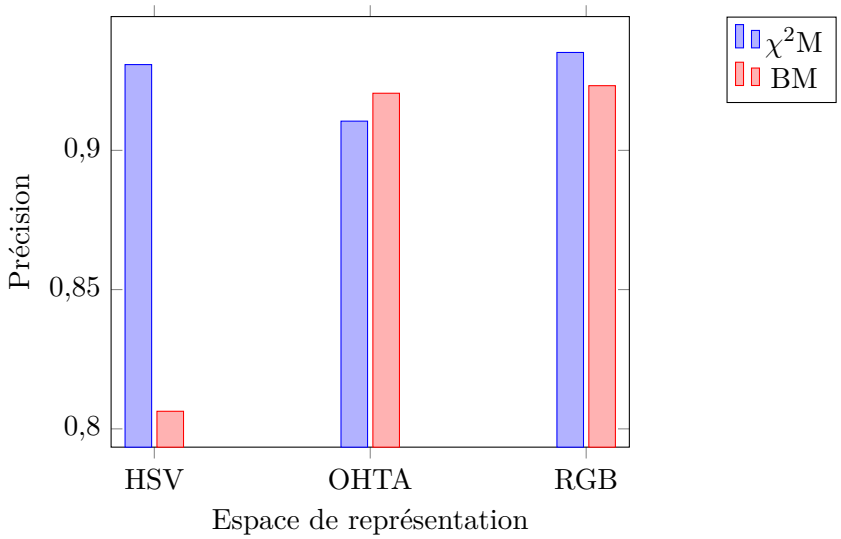

FIGURE 5.5 – Précision selon la mesure utilisée en fonction des espaces de représentation des couleurs différents et un nombre de partitions fixé à 350.

La Figure 5.5 présente la précision mesurée, en fonction de l'espace de couleur et de la mesure de similarité considérée, pour un nombre de partitions constant, fixé à 350. Nous remarquons que la combinaison Bhattacharyya-Mahalanobis produit une précision inférieure à la combinaison χ^2-Mahalanobis dans la plupart des cas. Cela est d'autant plus flagrant dans le cas de l'espace de couleur HSV qui voit la précision mesurée se dégrader drastiquement. Il n'y a que pour l'espace de représentation OHTA que la précision mesurée augmente. Cette dernière reste pour autant inférieure à celle mesurée sur l'espace de couleur RGB. Les partitions des histogrammes spatio-temporels sont plus homogènes avec l'espace de représentation OHTA du fait de l'absence de corrélation des canaux. La mesure χ^2 pondère les différences selon la taille des partitions, si les partitions sont homogènes, cela n'apporte rien. Cela peut expliquer pourquoi la distance de Bhattacharyya produit de meilleurs résultats par rapport à la mesure du χ^2.

Nous allons maintenant faire varier le nombre de partitions utilisées pour construire nos histogrammes spatio-temporels en utilisant l'espace de représentation RGB.

La Figure 5.6 montre que la précision de la mesure de similarité χ^2-Mahalanobis est plus élevé que celle de la mesure Bhattacharyya-Mahalanobis jusqu'au seuil de 1.500 partitions. À partir de ce seuil, la précision de la mesure χ^2-Mahalanobis se stabilise avant de commencer à diminuer. La précision de la mesure Bhattacharyya-Mahalanobis

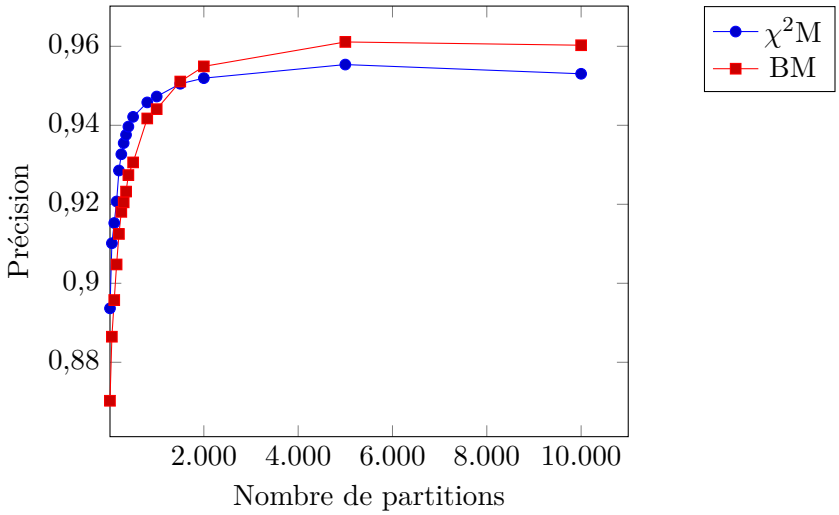

FIGURE 5.6 – Évolution de la précision en fonction du nombre de partitions des HST, entre 10 et 10.000 partitions, selon la mesure de similarité considérée.

continue de progresser jusqu'à atteindre un plafond autour de 5.000 partitions pour une précision d'environ 0,96.

Le comportement de l'évolution de la précision observée avec la mesure de similarité Bhattacharyya-Mahalanobis s'explique par le fait qu'un trop petit nombre de partitions fera diminuer la précision liée à la mesure de Bhattacharyya (cf Section 2.2.2) en surestimant la région de recouvrement. Un trop grand nombre de partitions fera diminuer la précision liée à la mesure de Bhattacharyya en créant des partitions vides de membres. De plus, l'information spatio-temporelle, mesurée par la distance de Mahalanobis, se retrouve de diluée dans plusieurs partitions quand le nombre de celles-ci devient trop grand. De ce fait, la précision diminue à partir d'un certain seuil. La mesure du χ^2 est moins sensible à cela. La mesure de similarité basée sur la distance de Bhattacharyya ne se compare favorablement à celle basée sur le χ^2 que quand le nombre de partitions est grand (> 1.500). On observe que pour un nombre de partitions plus faible, mis en évidence par la Figure 5.7, la distance basée sur le χ^2 est supérieure à l'autre. Cette différence de précision est maximale pour 50 partitions avec 2,4 points de précision de différence. Elle est quasiment nulle à 1.500 partitions. Néanmoins, la précision maximale atteinte avec la distance de Bhattacharyya n'est qu'un peu supérieure à celle atteinte avec la distance basée sur le χ^2.

5.5.5 Stratégie de construction

Dans la Section 4.3, nous avons proposé de plusieurs stratégies de construction pour les histogrammes spatio-temporels. Nous nous intéressons maintenant à la précision qu'il est possible d'atteindre avec chaque stratégie. Pour cela, nous comparons les résultats obtenus par les différentes stratégies de construction en utilisant l'espace de représentation RGB. Nous avons aussi fixé le nombre de partitions utilisées dans la construction à 350 pour les mêmes raisons que dans les expérimentations précédentes.

Nous comparons les précisions obtenues par la mise en correspondance des histo-

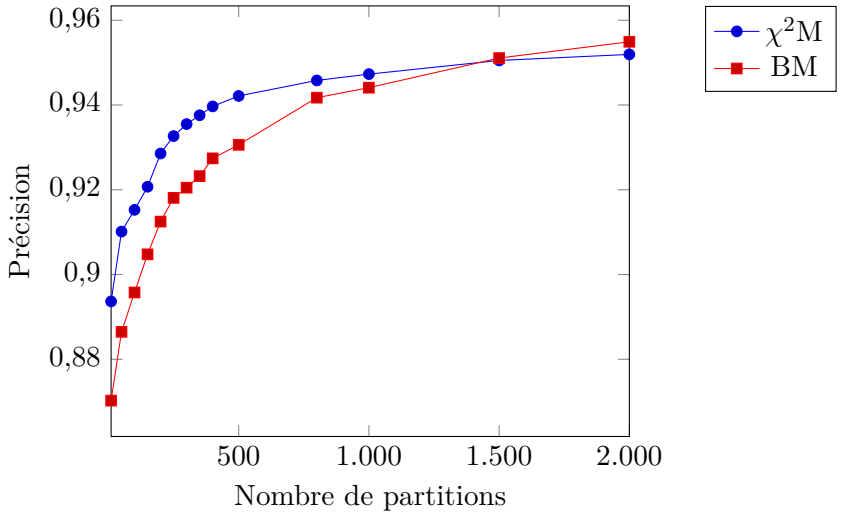

FIGURE 5.7 – Évolution de la précision en fonction du nombre de partitions des HST entre 10 et 2.000 partitions, selon la mesure de similarité considérée.

grammes spatio-temporels construits par cumul de l'information sur toutes les trames ("cm") avec celle obtenue par une fenêtre glissante ("slide") et une fenêtre sautante ("jump"). La mesure de similarité basée sur le χ^2 et la distance de Mahalanobis est utilisée pour comparer ces histogrammes spatio-temporels.

Enfin, la stratégie qui consiste à construire un histogramme spatio-temporel par canal de l'espace de couleur a été aussi comparé ("3d").

Pour les comparaisons où l'on a plus d'un histogramme spatio-temporel pour représenter une occurrence, la mesure de similarité est la même que celle proposée précédemment. Seule la mesure de similarité maximale obtenue en comparant chaque histogramme spatio-temporel d'une occurrence à tous les autres de l'autre occurrence, est conservée.

Comme le montre le diagramme de la Figure 5.8, la stratégie qui consiste à représenter toute l'occurrence vidéo d'une personne par un seul histogramme spatio-temporel est celle qui, de loin, donne les meilleurs résultats en termes de précision. La construction par fenêtre sautante donne des résultats inférieurs à ceux obtenus par la construction par fenêtre glissante. Pour rappel, le but de cette construction était de réduire le coût calculatoire de la construction par fenêtre glissante, en acceptant une perte de précision. La stratégie qui consiste à construire un histogramme spatio-temporel par canal de l'espace de couleur produit les résultats les plus faibles. Les canaux de l'espace de couleur RGB sont fortement corrélés. Le fait de séparer ces canaux dégrade l'information d'apparence. Cela explique que cette approche donne des résultats inférieurs aux autres. Il ressort de notre étude que les stratégies de construction d'histogrammes spatio-temporels par fenêtre ne permettent pas d'augmenter la précision. De plus, la construction des histogrammes spatio-temporels ainsi que leur comparaison est bien plus coûteuse à calculer (cf. Section 4.5). Bien que ces stratégies de construction puissent être intéressantes en termes de précision dans le cas de vidéos très longues, sans découpages en plans, le coût calculatoire devient rapidement prohibitif avec l'augmentation de la durée des vidéos. L'approche de construction par fenêtre est donc à éviter dans tous les cas.

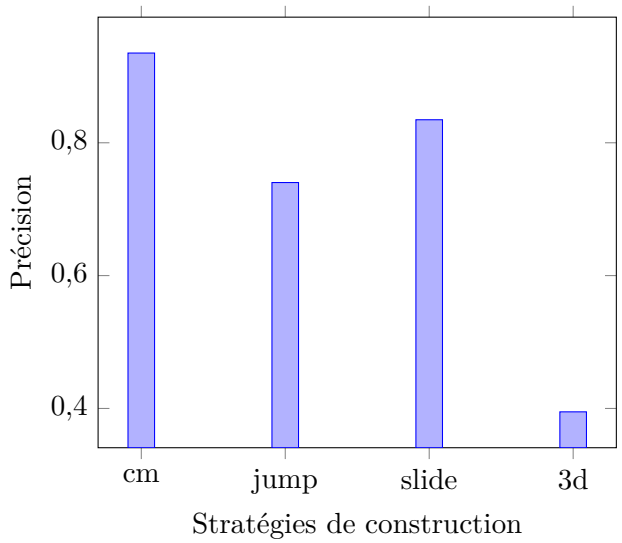

FIGURE 5.8 – Comparaison de la précision de plusieurs stratégies de construction de HST sur les OVP pour 350 partitions et l'espace de couleur RGB.

5.6 Significativité de la mesure de similarité

Afin de valider la pertinence de nos propositions et de nos résultats, nous avons réalisé un test statistique pour chacun d'entre eux. Ces tests nous permettent de vérifier que nous arrivons bien à discriminer les différentes occurrences de personnes différentes. En revanche, le test statistique ne nous permet pas de comparer les propositions.

5.6.1 Test de Student

Pour cela nous avons réalisé un test de Student basé sur les mesures de similarités obtenue sur le corpus présenté précédemment. Nous avons deux séries de mesures de similarités, une contenant les mesures de similarités obtenues entre des occurrences vidéo de personnes portant la même identité (S_1) et une autre obtenues à partir d'occurrences vidéo de personnes portant des identités différentes (S_2) tel que :

$$S_1 = \{s(hst_{o_i}, hst_{o_j}) \mid id(o_i) = id(o_j), i \neq j\} \quad (5.1)$$
$$S_2 = \{s(hst_{o_i}, hst_{o_j}) \mid id(o_i) \neq id(o_j), i \neq j\} \quad (5.2)$$

Notre test statistique permet de vérifier l'hypothèse nulle H_0 :

$$H_0 : \mu_1 = \mu_2 \quad (5.3)$$

où μ_1 est la moyenne de S_1 et μ_2 celle de S_2.

En d'autres termes, on part de l'hypothèse qu'il n'existe pas de différence significative entre les deux séries de données. Dans notre cas, cela revient à supposer que notre approche ne permet pas de discriminer les différentes personnes de la base REPERE.

Si la p-valeur (ou p-value) du test de Student est inférieure à 0,005 alors le test a rejeté l'hypothèse H_0 et aura démontré qu'il existe une différence entre les deux séries de données et que cette différence est significative.

La p-valeur indique la probabilité que les résultats obtenus soient dûs au hasard. Ainsi une p-valeur inférieure à 0,05 indique qu'il y a moins de 5% de chance que les résultats soit obtenus de façon aléatoire. Plus la p-valeur est faible plus la différence est significative. Le seuil de 0,05 a longtemps été estimé par la communauté scientifique comme étant suffisant. Récemment, ce seuil a fait débat dans la communauté [57]. Bien qu'un seuil de 5% soit suffisant pour déterminer la significativité des résultats, ce seuil ne permet pas de garantir la reproductibilité de l'expérimentation. Ainsi, de nombreux scientifiques souhaitent abaisser le seuil de significativité d'un facteur d'au moins 10. Nous utiliserons ainsi le seuil de 0,005 (0,5%), recommandé dans [57].

Plusieurs versions du test de Student existent, dont une qui suppose que les deux échantillons comparés ont la même variance, et une autre qui suppose que leurs variances sont différentes. Après avoir testé que les variances des deux échantillons étaient bien semblables (p-value égale à $2,2e^{-16}$), nous avons appliqué le test de Student correspondant sur toutes les données expérimentales.

5.6.2 Séries de données testées

Les différentes stratégies de constructions d'histogramme spatio-temporel ont été aussi vérifiées. Afin d'éviter tout biais, les similarités obtenues quand un élément est comparé à lui-même (valeur de 1) ont été retirées des séries de données S_1.

5.6.3 Significativité de la similarité

Il ressort que **toutes** les configurations mentionnées précédemment permettent de discriminer entre les personnes, la p-value de chaque test est à chaque fois très proche de zéro ($2,2e^{-16}$), ce qui correspond au meilleur score de significativité possible avec le test de Student implémenté dans R [3]. Nous avons aussi vérifié que dans chaque configuration, la moyenne des valeurs de S_1 était significativement plus grande que la moyenne des S_2. Encore une fois, **tous** les tests effectués vérifient cette propriété avec une p-valeur très proche de zéro.

5.6.4 Significativité de l'augmentation du nombre de partitions

Nous avons aussi voulu vérifier qu'augmenter le nombre de partitions d'un histogramme pour une configuration donnée augmente la moyenne des valeurs de S_1 et diminue la moyenne des valeurs de S_2.

Dans tous les tests la valeur moyenne de mise correspondance est significativement plus élevé. En revanche, la moyenne des valeurs de S_2 ne diminue pas de façon significative et cela pour tous les tests.

Ainsi, augmenter le nombre de partitions permet de mieux mettre en correspondance les personnes, mais ne permet pas d'améliorer le rejet. En d'autres termes, augmenter le nombre de partitions augmente la similarité d'occurrences de même identité, mais ne diminue pas la similarité entre occurrences d'identité différentes.

5.6.5 Résumé de la significativité de nos résultats

En conclusion, les différents tests statistiques effectués sur la mise en correspondance d'occurrences vidéo de personnes nous confirment que notre approche permet, de façon

3. The R Project for Statistical Computing : http ://www.r-project.org/

très significative, de discriminer les différentes personnes. Cela est vrai quel que soit le nombre de partitions et l'espace de couleur utilisés ou le type d'histogrammes utilisés (parmi les histogrammes spatio-temporels, les spatiogrammes et les histogrammes de couleurs) pour décrire les occurrences vidéo. La p-valeur des tests statistiques réalisés garantit à la fois la significativité de notre approche, ainsi que la reproductibilité de nos résultats. En revanche, les tests statistiques ne nous permettent pas de comparer la qualité des résultats entre les différentes approches.

5.7 Qualité du regroupement

Après avoir étudié les résultats en termes de précision des histogrammes spatio-temporels selon les paramètres utilisés, nous allons maintenant nous intéresser aux résultats obtenus lors du regroupement. Pour cela, les matrices de similarités calculées lors des expérimentations précédentes vont être utilisées afin de regrouper les occurrences vidéo d'une même personne. L'objectif est d'obtenir un seul groupe pour chaque identité. Nous allons ainsi évaluer la qualité du regroupement en utilisant différents indices appropriés. Chacun permet d'évaluer un aspect particulier du regroupement.

5.7.1 Regroupement hiérarchique ascendant

L'approche utilisée pour effectuer le regroupement est un clustering hiérarchique ascendant que nous avons présenté dans la Section 4.8. Le regroupement est initialisé en mettant dans un même groupe les occurrences vidéo de personnes avec une similarité supérieure ou égale à 0,99. Pour rappel, cette valeur nous permet de mettre en correspondance deux occurrences vidéo sans faire d'erreur. Pour plus de détails à ce sujet sont présentés dans la Section 4.8.

Nous avons appliqué cette méthode de regroupement sur des histogrammes spatio-temporels construits sur différents espaces de couleurs, avec un nombre différent de partitions et des mesures de similarités différentes. Cela permet, en outre, de vérifier si une configuration avec une précision plus faible dans une tâche de recherche offre les mêmes performances lors du regroupement d'occurrences.

Nous notons Ω notre regroupement, il s'agit d'un ensemble de groupes ω contenant des occurrences vidéo de personnes o tel que :

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\} \tag{5.4}$$

et

$$\omega_i = \{o_1^i, o_2^i, \ldots, o_{|\omega_i|}^i\} \tag{5.5}$$

Dans notre regroupement, pour l'évaluation, on s'intéresse au sous-ensemble C des identités $\mathbb{I}$ présentes dans celui-ci :

$$C = \{\iota_i \in \mathbb{I} \mid \exists o \in \mathbb{O}, id(o) = \iota_i\} \tag{5.6}$$

5.7.2 Mesure de pureté

La pureté [48] est une mesure de l'homogénéité des groupes. Autrement dit, cette mesure vérifie si les éléments au sein d'un groupe appartiennent à une même classe. Pour calculer la pureté, on commence par attribuer à chaque groupe l'étiquette de l'identité la

plus fréquente parmi ses membres. Ensuite, il s'agit d'un simple calcul de précision dont voici la formule :

$$\text{purete}(\Omega, C) = \frac{1}{\sum_{k=0}^{|\Omega|-1} |\omega_k|} \sum_{k=0}^{|\Omega|-1} \max_j |\omega_k \cap id^{-1}(\iota_j)| \tag{5.7}$$

Comme toutes les occurrences vidéo font partie du regroupement, cette définition peut être simplifié en :

$$\text{purete}(\Omega, C) = \frac{1}{|\mathbb{O}|} \sum_{k=0}^{|\Omega|-1} \max_j |\omega_k \cap id^{-1}(\iota_j)| \tag{5.8}$$

Le cas particulier où plusieurs classes auraient la même fréquence n'influence pas le résultat : on obtient le même résultat quel que soit l'étiquette sélectionnée. Par ailleurs, on peut noter qu'une pureté parfaite de 1 peut être obtenue si chaque élément est dans un cluster différent (i.e. il y a autant de clusters que d'éléments).

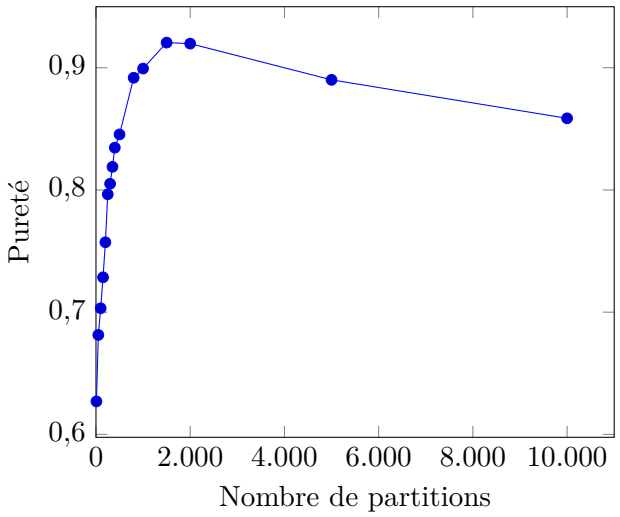

FIGURE 5.9 – Pureté du regroupement, en fonction du nombre de partitions compris entre 10 et 10.000 de l'histogramme spatio-temporel sur l'espace de couleur RGB.

On observe que la pureté de notre regroupement, pour un nombre relativement faible de partitions, Figures 5.9 et 5.10, augmente avec le nombre de partitions. Cela reste vrai jusqu'à un seuil légèrement inférieur à 2.000 partitions, à partir duquel la pureté diminue graduellement.

Ce comportement suit celui observé lors du calcul de la précision dans une tâche de recherche que nous avons présenté précédemment. En effet, nous remarquons qu'à partir d'un certain seuil (au-delà de 1.500 partitions), le trop grand nombre de partitions dilue l'information et ne permet plus, de façon aussi efficace, de discriminer entre les personnes.

5.7.3 Mesure de fragmentation

La fragmentation quantifie la dispersion de chaque identité dans différents clusters. La formule de la fragmentation est [48] :

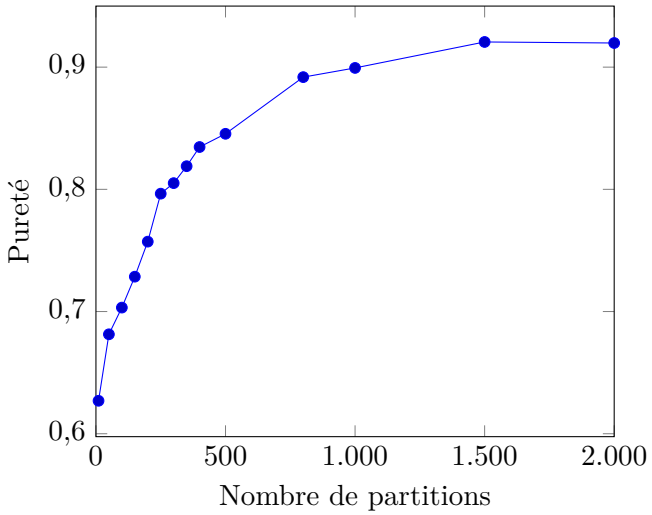

FIGURE 5.10 – Pureté du regroupement, en fonction du nombre de partitions compris entre 10 et 2.000 de l'histogramme spatio-temporel sur l'espace de couleur RGB.

$$frag(\Omega, C) = \frac{\sum_{i=0}^{|C|-1} |\{\omega \in \Omega | \exists o \in id^{-1}(\iota_i), o \in \omega\}|}{|C|} \tag{5.9}$$

La fragmentation mesure, en moyenne, dans combien de groupes apparaît chaque identité du regroupement. Ainsi, si chaque identité est représentée par un cluster dédié, la fragmentation est de 1. Il est important de noter qu'une fragmentation inférieure à 1 est obtenue si le nombre de clusters est inférieur au nombre d'identités.

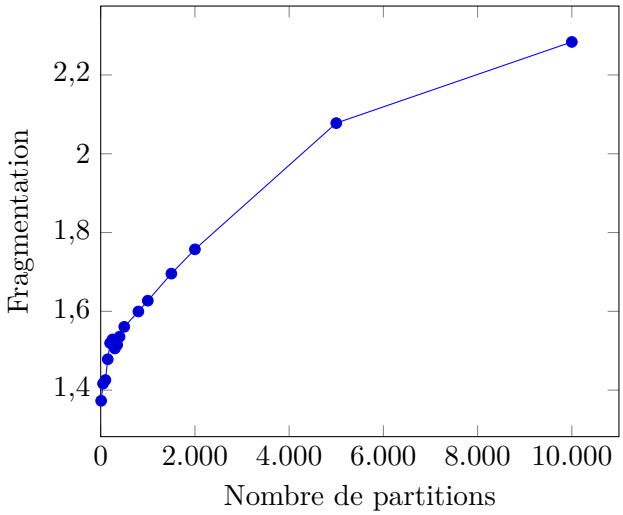

FIGURE 5.11 – Fragmentation en fonction du nombre de partitions compris entre 10 et 10.000 de l'histogramme spatio-temporel sur l'espace de couleur RGB.

On observe dans notre regroupement que l'indice de fragmentation (cf. Figures 5.11 et 5.12) diminue en augmentant le nombre de partition. Ceci indique que chaque identité

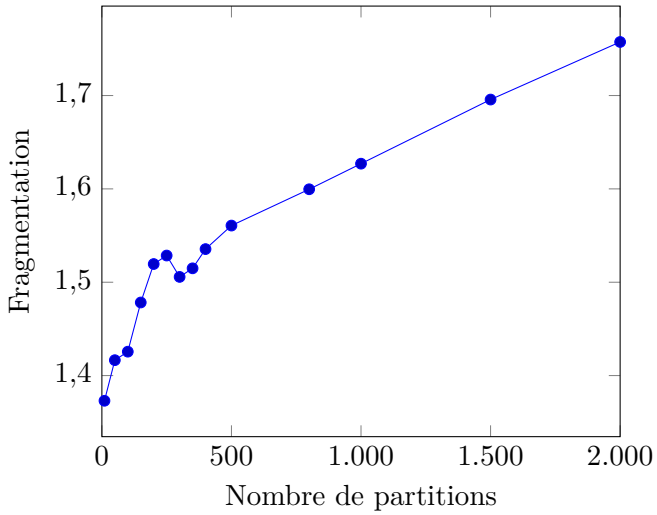

FIGURE 5.12 – Fragmentation en fonction du nombre de partitions compris entre 10 et 2.000 de l'histogramme spatio-temporel sur l'espace de couleur RGB.

est représentée par plusieurs clusters. Ainsi, de plus en plus de clusters sont créés avec l'augmentation du nombre de partitions des histogrammes spatio-temporels.

Nous interprétons cela par le fait qu'en augmentant le nombre de partitions, la mesure de similarité devient moins tolérante. Nous avons observé lors des tests statistiques que la mesure de similarité, entre les occurrences associées à des identités différentes, ne diminue pas avec l'augmentation du nombre de partitions. Cependant, la similarité des occurrences de même identité augmente. Ainsi, les éléments étant en moyenne plus similaires entre eux, de moins en moins de clusters peuvent être regroupés par le clustering hiérarchique. Cela est d'ailleurs confirmé par la pureté, vue précédemment, qui augmente avec le nombre de partitions (jusqu'à un certain seuil).

5.7.4 Vrais/faux positifs/négatifs

Pour évaluer un regroupement, il est courant de se baser sur le nombre de "vrais positifs" (TP), "vrais négatifs" (TN), "faux positifs" (FP) et "faux négatifs" (FN). Ceux-ci sont définis de la façon suivante :

$$TP = |\{(o,o') \in \mathbb{O}^2 | id(o) = id(o'), o \in \omega_i, o' \in \omega_i\}| \quad (5.10)$$

$$TN = |\{(o,o') \in \mathbb{O}^2 | id(o) \neq id(o'), o \in \omega_i, o' \in \omega_j, i \neq j\}| \quad (5.11)$$

$$FP = |\{(o,o') \in \mathbb{O}^2 | id(o) \neq id(o'), o \in \omega_i, o' \in \omega_i\}| \quad (5.12)$$

$$FN = |\{(o,o') \in \mathbb{O}^2 | id(o) = id(o'), o \in \omega_i, o' \in \omega_j, i \neq j\}| \quad (5.13)$$

$TP + TN$ peut être vu comme le nombre d'occurrences correctement regroupées et $FP + FN$ comme le nombre d'erreurs de regroupement commises.

5.7.5 Indice de Rand

L'indice de Rand [91] est une mesure de similarité entre deux partitions d'un ensemble. Il est principalement utilisé en catégorisation automatique. Son principe est, pour chaque paire d'objets, de voir si elle a été classée de la même façon (ensemble ou séparément) dans les deux partitions.

L'indice de Rand (RAND) se défini de la façon suivante :

$$\text{RAND} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.14}$$

L'indice de Rand mesure d'une certaine façon le taux de réussite du clustering. En effet, le nombre de classifications correctes ($TP + TN$) est divisé par le nombre total de classifications.

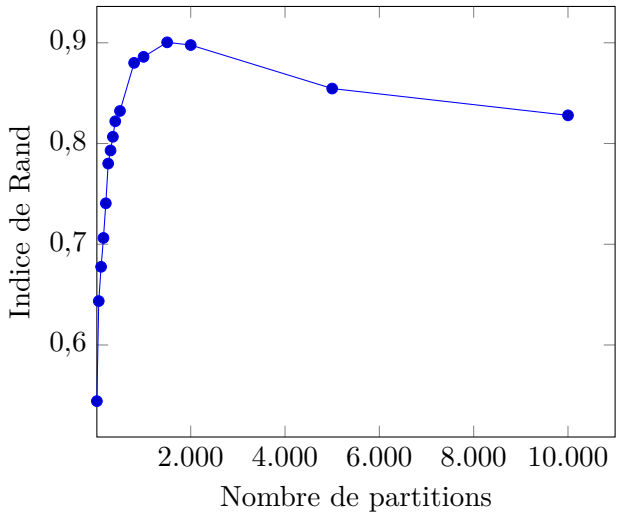

FIGURE 5.13 – Indice de Rand en fonction du nombre de partitions, compris entre 10 et 10.000, de l'histogramme spatio-temporel construit sur l'espace de couleur RGB.

On observe, dans notre regroupement, Figures 5.13 et 5.14, que l'indice de Rand progresse fortement en augmentant le nombre de partitions. Cette progression s'arrête autour de 2.000 partitions où l'indice de Rand entame une diminution progressive. Cette évolution correspond à celle observée précédemment lors du calcul de la pureté ou encore de la précision dans une tâche de recherche.

5.7.6 Rappel/Précision

Les deux mesures les plus emblématiques de l'évaluation de résultats sont probablement la précision et le rappel. La précision est le nombre de résultats pertinents retrouvés rapporté au nombre de résultats total proposés. Le rappel est défini par le nombre de résultats pertinents retrouvés au regard du nombre de résultats pertinents possibles. La précision P et le rappel R sont définis de la façon suivante :

$$P = \frac{TP}{TP + FP} \tag{5.15}$$

$$R = \frac{TP}{TP + FN} \tag{5.16}$$

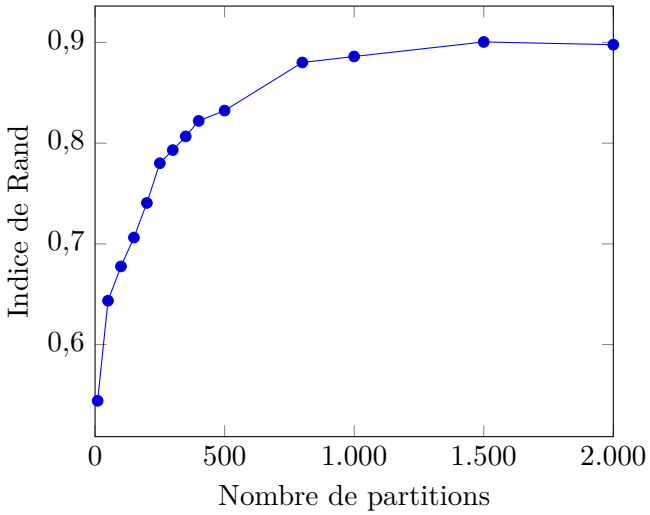

FIGURE 5.14 – Indice de Rand en fonction du nombre de partitions, compris entre 10 et 2.000, de l'histogramme spatio-temporel construit sur l'espace de couleur RGB.

5.7.7 F-mesure

La F-mesure [76] est utilisée pour équilibrer la contribution des faux négatifs à la mesure en pondérant le rappel à travers un paramètre $\beta \geq 0$. Cela nous permet de calculer la F-mesure en utilisant la formule suivante :

$$F_\beta = \frac{(\beta^2 + 1) * P * R}{\beta^2 * P + R} \tag{5.17}$$

Notez que quand $\beta = 0$, $F_0 = P$. En d'autres terme, le rappel n'a aucun impact sur la F-mesure quand $\beta = 0$. Augmenter β confère un poids de plus en plus important au rappel dans la F-mesure. Dans notre regroupement, on étudie l'évolution de la F1-Mesure en fonction du nombre de partitions. Cette mesure prend en compte la précision et le rappel sans pondération particulière entre les deux indices.

On observe, dans les Figures 5.15 et 5.16, que la mesure F1 augmente avec le nombre de partitions jusqu'au seuil de 2.000 partitions où elle entame une diminution progressive. L'évolution de la mesure F1 correspond en tous points à celle observé avec les autres indices.

5.7.8 Indice Fowlkes–Mallows

L'indice de Fowlkes–Mallows [37] est une méthode d'évaluation externe qui est utilisée pour comparer la similarité entre deux regroupements. Cette mesure de similarité peut être soit entre deux clustering hiérarchique soit entre un clustering et une classification servant de vérité terrain. Une valeur élevée de l'indice Fowlkes–Mallows indique une similarité élevée entre les deux regroupements.

L'indice Fowlkes–Mallows (FM) est défini de la façon suivante :

$$\text{FM} = \sqrt{\frac{TP}{TP + FP} * \frac{TP}{TP + FN}} \tag{5.18}$$

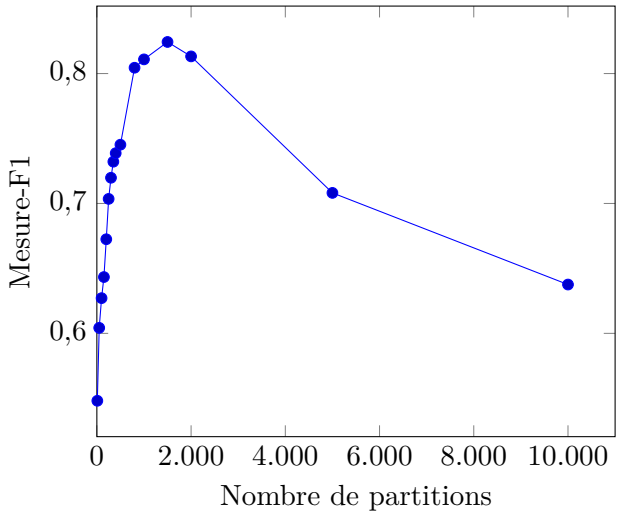

FIGURE 5.15 – Mesure F1 en fonction du nombre de partitions, compris entre 10 et 10.000, de l'histogramme spatio-temporel sur l'espace de couleur RGB.

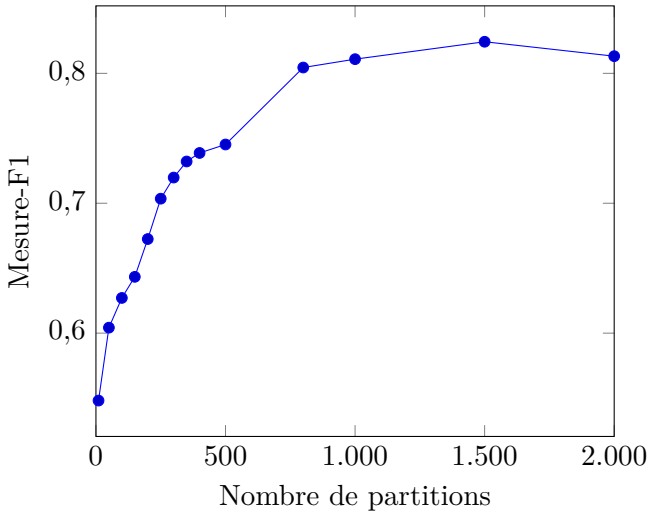

FIGURE 5.16 – Mesure F1 en fonction du nombre de partitions, compris entre 10 et 2.000, de l'histogramme spatio-temporel sur l'espace de couleur RGB.

Où TP est le nombre de vrais positifs, FP est le nombre de faux positifs et FN est le nombre de faux négatifs. Cette équation peut aussi s'écrire, plus simplement :

$$\mathrm{FM} = \sqrt{P * R} \tag{5.19}$$

Où P est la précision et R le rappel.

L'évolution de l'indice de Fowlkes-Mallows en fonction du nombre de partitions, Figures 5.17 et 5.18 est similaire à l'évolution des autres indices où de la précision dans une tâche de recherche.

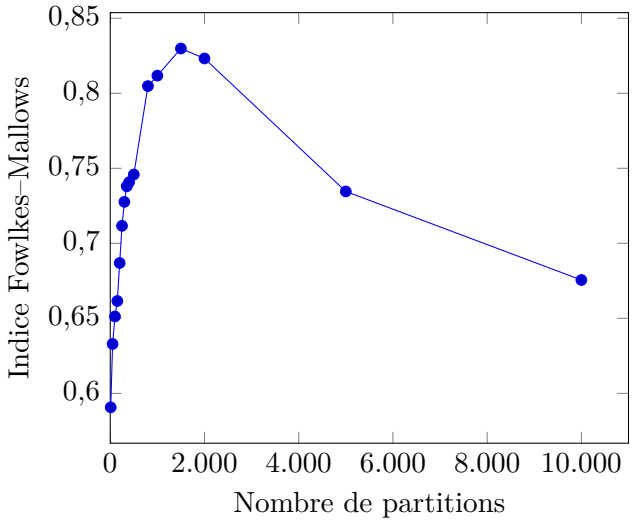

FIGURE 5.17 – Indice Fowlkes-Mallows en fonction du nombre de partitions, compris entre 10 et 10.000, de l'histogramme spatio-temporel sur l'espace de couleur RGB.

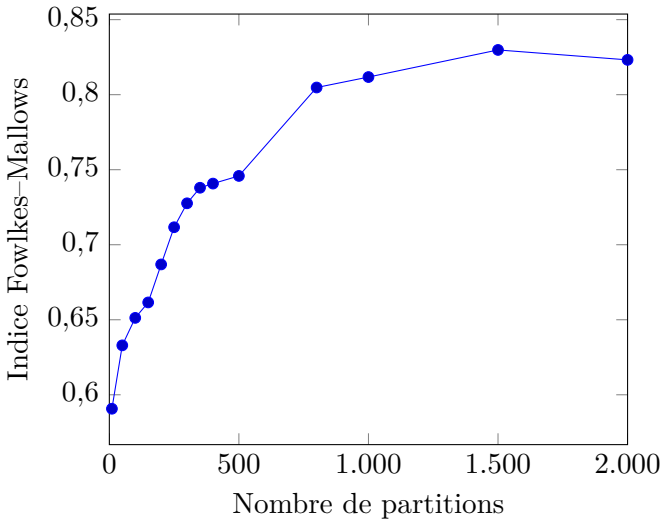

FIGURE 5.18 – Indice Fowlkes-Mallows en fonction du nombre de partitions, compris entre 10 et 2.000, de l'histogramme spatio-temporel sur l'espace de couleur RGB.

5.7.9 Résumé de la qualité du regroupement

Dans les différents indices mesurant la qualité du regroupement, nous observons une rapide progression entre 10 et 2.000 partitions avant une diminution progressive. Nous expliquons ce comportement par le fait que dans un premier temps, augmenter le nombre de partitions permet de mieux caractériser l'information spatio-temporelle. Les informations non corrélées sont rangées dans une même partition ce qui renforce le pouvoir descriptif de l'ensemble. Quand le nombre de partitions commence à être trop grand, les informations commencent à être séparées dans des partitions différentes. Les histogrammes spatio-

temporels deviennent trop discriminants et ne permettent plus de mesurer la similarité entre deux occurrences vidéo d'une même personne dans des conditions trop différentes. Ainsi, au-delà d'un certain seuil, augmenter le nombre de partitions donne des résultats inférieurs tout en augmentant la quantité de calculs à réaliser. Il est ainsi important d'identifier ce seuil dans les différentes applications où les histogrammes spatio-temporels interviennent.

Dans notre application, sur des émissions audiovisuelles, nous avons observé que ce seuil se situe à 1.500 partitions. Nous supposons que dans différentes applications ce seuil devrait varier, bien que de manière limitée car notre corpus propose des émissions très différentes. Ainsi, il peut être intéressant de chercher ce seuil, par exemple par une approche dichotomique, entre 1.000 et 2.000 partitions.

5.8 Précision et clustering

Nous avons remarqué que l'évolution de la précision et des différents indices de qualité du regroupement sont semblables. Nous avons voulu comparer la précision avec ces différents indices pour vérifier s'il était possible d'estimer les propriétés du clustering à partir du calcul de la précision à n_i sur la matrice de similarités.

Pour comparer la précision aux différents indices, nous les avons affichés l'un en fonction de l'autre et avons calculé une droite de régression linéaire.

5.8.1 Précision et pureté

Dans un premier temps nous comparons la précision à la pureté en affichant la pureté en fonction de la précision.

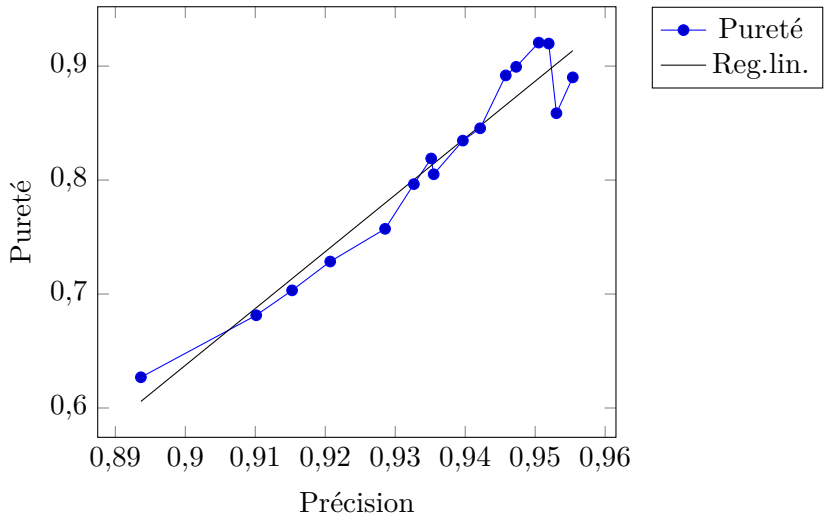

FIGURE 5.19 – Pureté du clustering en fonction de la précision mesurée.

Dans la Figure 5.19, la courbe présentant la pureté en fonction de la précision suit de près la droite de régression linéaire. On peut donc en conclure que dans notre application aux émissions audiovisuelles, la pureté du clustering peut être correctement estimée à

partir de la précision. Il est ainsi possible de fixer le paramétrage des histogrammes spatio-temporels afin de répondre à un objectif de pureté du regroupement de personnes.

5.8.2 Précision et fragmentation

D'une façon similaire, nous avons voulu comparer la précision à la fragmentation du clustering. Dans nos expérimentations, nous avons observé que la fragmentation était croissante avec l'augmentation du nombre de partitions.

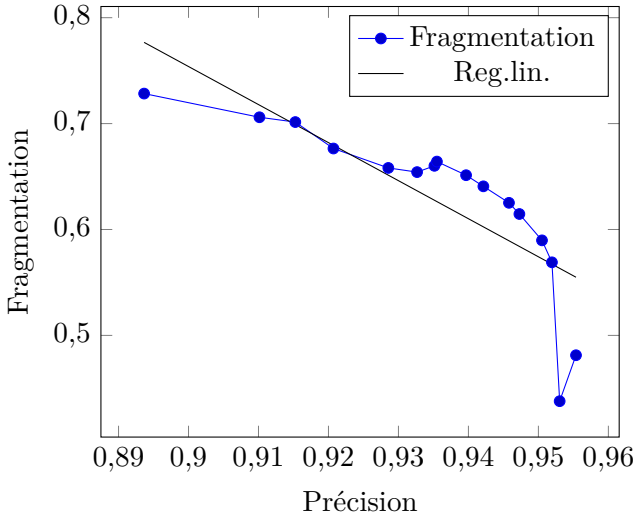

FIGURE 5.20 – Fragmentation du clustering en fonction de la précision mesurée.

Cependant, quand nous affichons la fragmentation en fonction de la précision Figure 5.20 nous observons que la courbe présentant la fragmentation en fonction de la précision est loin de former une droite. De ce fait, elle ne peut pas être estimée correctement par une droite de régression linéaire. Ainsi, la fragmentation peut difficilement être estimée à partir du taux de précision mesuré.

5.8.3 Précision et Rand

Nous avons enfin voulu voir si la précision permettait d'estimer l'indice de Rand. Pour cela, nous avons affiché l'indice de Rand en fonction de la précision et calculé et affiché la droite de régression linéaire.

Nous observons dans la Figure 5.21 que la courbe présentant l'indice de Rand en fonction de la précision suit de très près la droite de régression linéaire. L'indice de Rand peut donc être estimé de façon relativement précise à partir du taux de précision mesuré pour choisir les paramètres des histogrammes spatio-temporels.

5.8.4 Résumé de la mesure de précision pour l'évaluation du clustering

Nous avons vu que, dans la plupart des cas, les indices d'évaluation du clustering pouvaient être estimés à partir de la précision à n d'une tâche de recherche. Ce résultat est particulièrement utile à prendre en considération lors du paramétrage des histogrammes spatio-temporels. Il est ainsi facile de prédire les qualités du regroupement à partir de

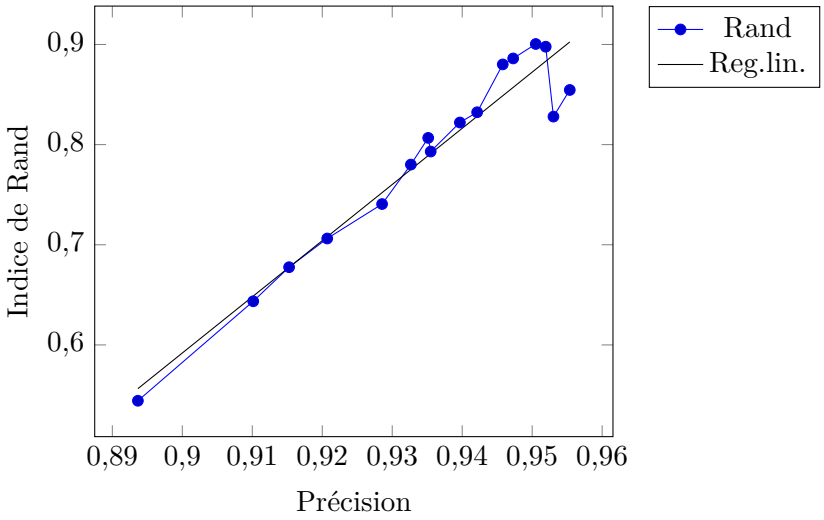

FIGURE 5.21 – Indice de Rand du clustering en fonction de la précision mesurée.

la simple mesure de précision. Il n'y a que la fragmentation qui semble difficile à prédire simplement.

Dans ces conditions, il est possible de se fixer des objectifs de qualité du clustering. Inversement, il est possible de respecter des contraintes de coût calculatoire tout en prédisant l'impact de celles-ci sur le regroupement.

5.9 Comparaison avec des méthodes existantes

Après avoir étudié le fonctionnement des histogrammes spatio-temporels nous allons les comparer avec d'autres descripteurs de l'état de l'art en termes de performances et de de coût calculatoire. Pour cela, nous comparons les résultats obtenus par les histogrammes spatio-temporels à ceux obtenus par les histogrammes de couleur et les spatiogrammes qui sont les principales approches concurrentes à la nôtre.

5.9.1 Précision

Dans un premier temps, nous comparons l'évolution de la précision des différentes approches en faisant varier le nombre de partitions. Cela nous permet de vérifier si les différentes approches évoluent de la même façon. De plus, cela nous permet de comparer les résultats en termes de précision entre les différentes approches.

On remarque dans les Figures 5.22 et 5.23 que la précision évolue de façon parallèle entre les différentes approches. La précision des différentes approches progresse rapidement pour un petit nombre croissant de partitions (entre 10 et 1.000). La précision atteint un maximum pour commencer à diminuer à partir de 5.000 partitions. Notre approche basée sur les histogrammes spatio-temporels obtient une meilleure précision que les approches basées sur les histogrammes de couleur ou les spatiogrammes. Ceci confirme notre hypothèse selon laquelle l'information spatio-temporelle est importante pour distinguer les occurrences vidéo de personnes.

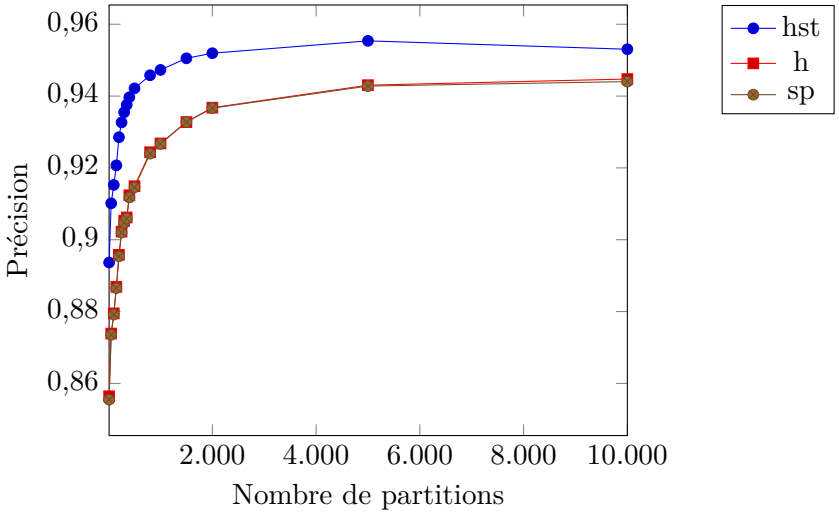

FIGURE 5.22 – Évolution de la précision en fonction du nombre de partitions du modèle entre 10 et 10.000 partitions. Les courbes correspondants à l'histogramme et au spatiogramme se chevauchent ici.

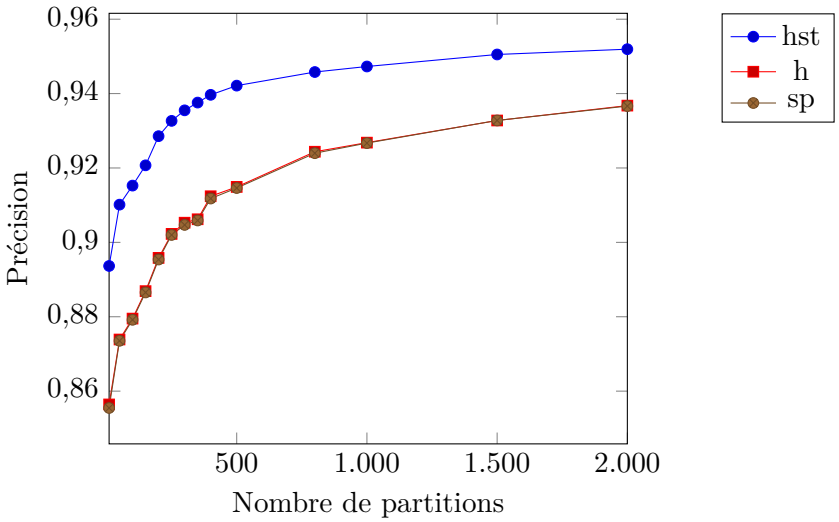

FIGURE 5.23 – Évolution de la précision en fonction du nombre de partitions du modèle entre 10 et 2.000 partitions. Les courbes correspondants à l'histogramme et au spatiogramme se chevauchent ici.

Il est intéressant d'observer que la précision des spatiogrammes est quasiment identique à celle des histogrammes de couleur. Il semble ainsi que l'information spatiale seule ne soit pas pertinente pour discriminer entre les occurrences vidéo de personnes et que la couleur seule donne de bons résultats.

5.9.2 Efficience

Après avoir comparé la précision de chaque approche, nous allons comparer leur efficience, pour cela nous avons calculé l'efficience relative de chaque approche. Celle-ci se calcule en divisant le gain en précision par le gain en nombre de partitions entre deux mesures successives. Elle compare ainsi le gain en précision en fonction du gain en complexité, représenté ici par l'augmentation du nombre de partitions. Cela permet de mesurer jusqu'à quel point il est utile d'augmenter le nombre de partitions de chaque approche.

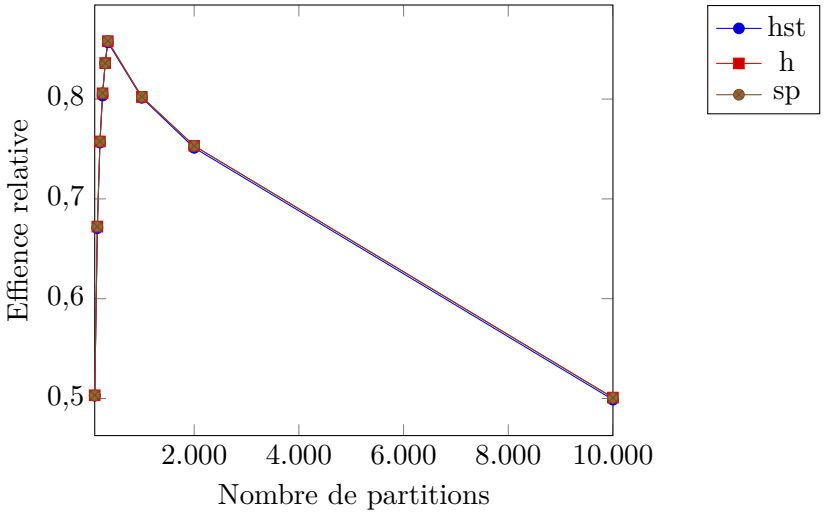

FIGURE 5.24 – Efficience relative de la précision par rapport à l'augmentation entre 10 et 10.000 du nombre de partitions utilisées pour la construction. Les trois courbes se chevauchent ici.

Dans les Figures 5.24 et 5.25, on observe que l'efficience relative des différentes approches est identique et est croissante jusqu'à un seuil de partitions de 500, à partir de ce point, l'efficience commence à diminuer de plus en plus rapidement. Cela signifie que pour un nombre de partitions supérieur à 500, la complexité augmente plus rapidement que la précision. Ainsi, à partir de 500 partitions, il est de moins en moins intéressant d'augmenter ce nombre de partitions. Cependant, dépasser ce seuil permet d'augmenter la précision mais chaque gain en précision à un coût de plus en plus élevé. Ceci est mis en évidence par le calcul de l'efficience absolue qui regarde le coût total de la précision.

Les Figures 5.26 et 5.27 montrent l'efficience absolue de la précision par rapport au nombre de partitions. On remarque de nouveau que les trois approches comparées ont une efficience absolue quasiment identique. Cette efficience absolue est rapidement décroissante entre 10 et 2.000 partitions, quelque soit l'approche considérée et continue de diminuer, mais moins rapidement à partir de ce point. L'efficience absolue est très faible à partir de 5.000 partitions et peut être considérée comme nulle à 10.000 partitions.

Cette efficience absolue nous indique qu'il n'est pas utile d'utiliser plus de 2.000 partitions, et que pour un nombre supérieur de partitions, le système perd complètement en efficience. On remarque en effet que la précision des différentes approches cesse d'augmenter à partir de ce seuil pour, au contraire, perdre en précision (voir Figure 5.22). Dans notre application aux émissions audiovisuelles, un nombre de partitions compris autour

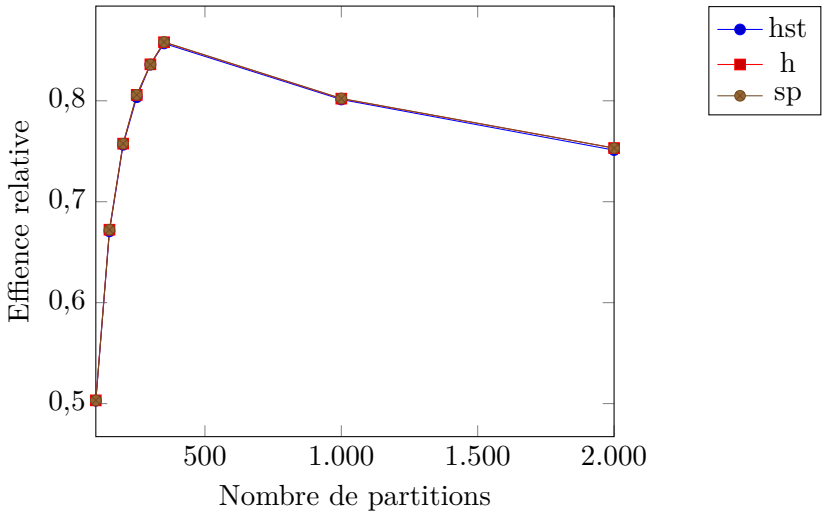

FIGURE 5.25 – Efficience relative de la précision par rapport à l'augmentation entre 100 et 2.000 du nombre de partitions utilisées pour la construction. Les trois courbes se chevauchent ici.

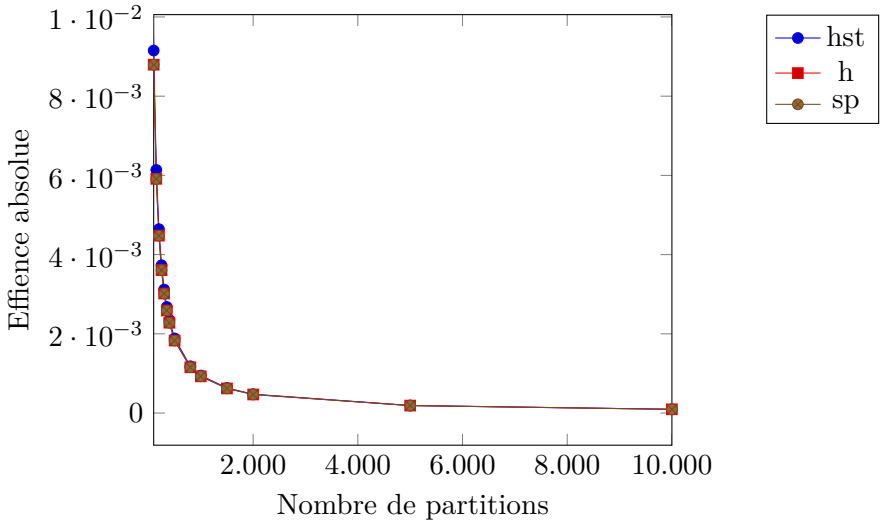

FIGURE 5.26 – Efficience absolue de la précision par rapport au nombre de partitions, compris entre 100 et 10.000, utilisées pour la construction. Les trois courbes se chevauchent ici.

de 1.500, pour les histogrammes spatio-temporels, permet d'atteindre un bon rapport précision/coût calculatoire.

5.9.3 Précision à coût mémoire constant

Afin de confirmer la pertinence de notre approche, nous avons voulu comparer les différentes approches en fonction de leur coût mémoire. Comme nous l'avons étudié dans

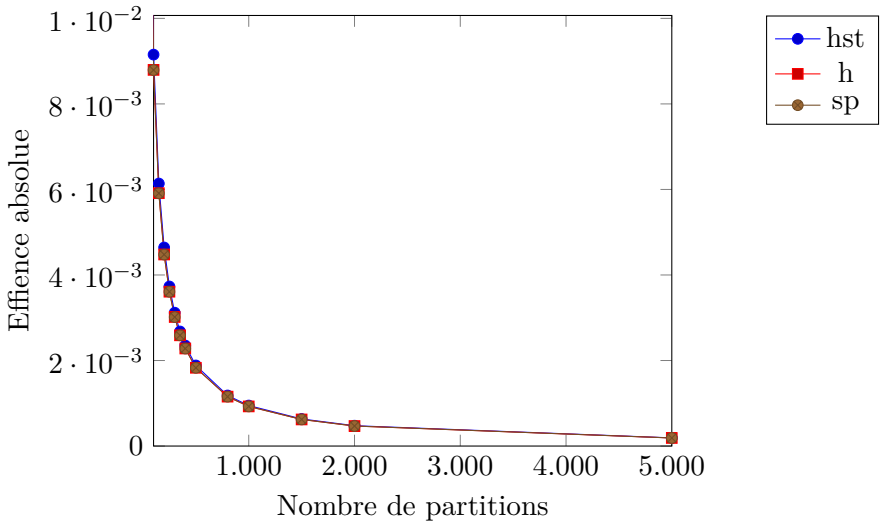

FIGURE 5.27 – Efficience absolue de la précision par rapport au nombre partitions, compris entre 100 et 5.000, utilisées pour la construction. Les trois courbes se chevauchent ici.

la Section 4.5, les histogrammes spatio-temporels, les spatiogrammes et les histogrammes de couleurs construit par cumul ont des complexités similaires pour leur construction et leur comparaison. Seul le coût mémoire occupé par ces descripteurs varie de façon significative. Ainsi, nous allons comparer la précision des différentes approches en prenant en compte cette différence de coût mémoire.

Nous utilisons comme base les histogrammes de couleurs avec un coût mémoire de 1 par partition (la donnée de comptage). Pour rappel, les spatiogrammes ont un coût mémoire de 6 par partition qui comprend les données de comptage, les positions moyennes $\bar{x}$, $\bar{y}$ et les covariances $cov(x,x)$, $cov(x,y)$ et $cov(y,y)$. Les histogrammes spatio-temporels ont un coût mémoire de 9 par partition qui comprend les données des spatiogrammes en ajoutant la position moyenne $\bar{t}$ ainsi que les covariances associées au temps $cov(x,t)$, $cov(y,t)$ et $cov(t,t)$.

Les Figures 5.28 et 5.29 montre la précision des différentes approches en fonction du coût mémoire relatif aux histogrammes de couleur. On remarque que pour un coût mémoire inférieur à 4.500, les histogrammes de couleurs donnent la meilleure précision, bien qu'ils soient progressivement rattrapés par les histogrammes spatio-temporels. Pour un coût de 4.500, les histogrammes spatio-temporels et les histogrammes de couleurs ont des précisions équivalentes. Cela signifie qu'un histogramme spatio-temporel de 500 partitions est équivalent à un histogramme de couleurs de 4.500 partitions. Au-delà d'un coût mémoire de 4.500, les histogrammes spatio-temporels ont une précision croissante avec le coût mémoire, alors que la précision des histogrammes de couleurs diminue. Ainsi, il n'existe pas d'histogrammes de couleurs pouvant atteindre la précision des histogrammes spatio-temporels quand ceux-ci possèdent plus de 500 partitions.

Les spatiogrammes sont en retrait en termes de précision en fonction du coût mémoire relatif à celui des histogrammes de couleur. Ce n'est qu'à partir d'un coût mémoire de 30.000 que la précision des spatiogrammes arrive à égaler puis à dépasser celle des histogrammes de couleurs. Ainsi, un spatiogramme de 5.000 partitions est équivalent

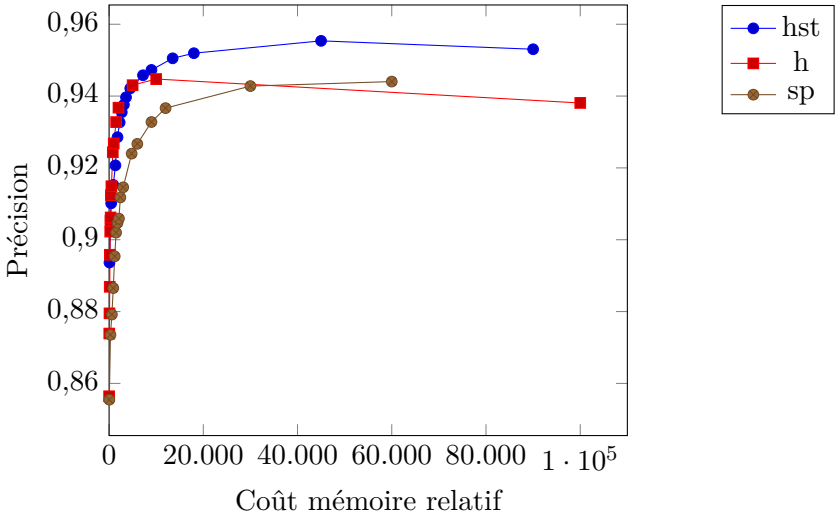

FIGURE 5.28 – Évolution de la précision en fonction du coût mémoire, relatif aux histogrammes de couleur.

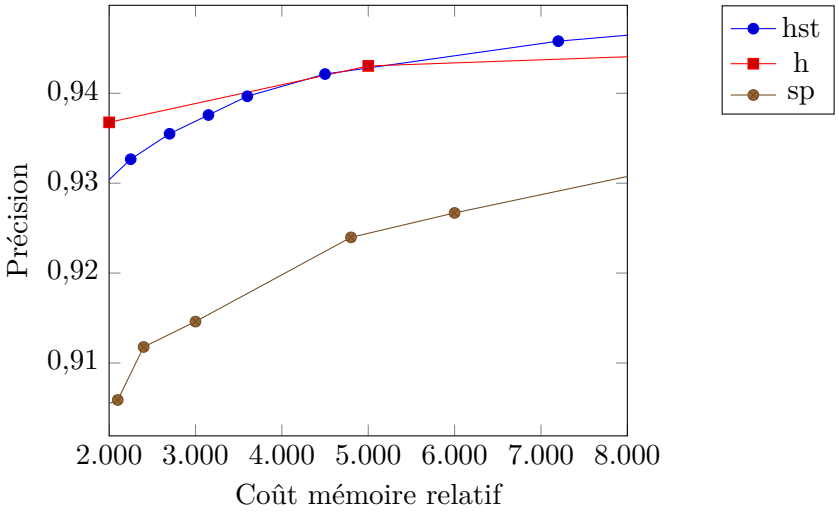

FIGURE 5.29 – Évolution de la précision en fonction du coût mémoire, relatif aux histogrammes de couleur.

à un histogramme de couleurs de 30.000 partitions. La précision des spatiogrammes ne semble pas rattraper celle des histogrammes spatio-temporels, elles semblent évoluer parallèlement l'une à l'autre.

En résumé, les histogrammes spatio-temporels surpassent, pour des coûts mémoire équivalents, les performances des histogrammes de couleurs et des spatiogrammes. Ces résultats mettent bien en évidence la contribution de la composante temporelle prise en compte dans les histogrammes spatio-temporels pour mettre en correspondance des occurrences vidéo de personnes. Cela valide donc notre hypothèse que l'aspect temporel des vidéos est porteur d'une information qui, couplée à l'information spatiale, permet

de distinguer les personnes. De plus, cela valide de façon expérimentale notre approche. Il est également important de noter que l'information spatiale seule ne permet pas de mettre en correspondance des occurrences vidéo de personnes de façon plus efficace qu'en utilisant de simples histogrammes de couleurs.

5.10 Résumé des résultats des expérimentations

Nous avons proposé une approche qui consiste à mettre en correspondance des occurrences vidéo de personnes afin de les regrouper par personnes. Dans les expérimentations, nous avons validé expérimentalement chaque étape de notre approche (présentée dans le Chapitre 4. Pour cela, nous avons appliqué notre approche à des émissions audiovisuelles réelles, issues de BFMTV et LCP.

De plus, nous avons testé, les paramètres optimaux des histogrammes spatio-temporels. Cela nous a permis de montrer que l'espace de couleur RGB permet de mieux discriminer les occurrences vidéo de personnes que les autres espaces de couleurs testés. En concordance avec nos hypothèses, il est effectivement possible de prédire les propriétés du regroupement obtenu à partir de la précision mesurée lors du paramétrage des histogrammes spatio-temporels. Il est ainsi possible de fixer des objectifs en termes de qualité du regroupement et de choisir de cette façon les paramètres. Inversement, il est possible de fixer la complexité du système et de prédire l'impact que cela aura lors du regroupement de personnes.

Enfin, nous avons comparé notre approche à d'autres approches basées sur les histogrammes de couleurs ou les spatiogrammes. Ainsi, à complexité égale, les histogrammes spatio-temporels permettent d'obtenir de meilleurs résultats que les spatiogrammes ou que les histogrammes de couleurs. Ceci confirme que la composante temps complémente de façon très significative la composante spatiale pour mettre en correspondance des occurrences vidéo de personnes et donc de les regrouper. Cette contribution est d'autant plus importante quand le nombre de classes est petit, avec plus de 200 points de base de précision gagnés, par rapport à d'autres approches utilisant pourtant du cumul d'information, pour les plus petits nombre de classes. De plus, les histogrammes spatio-temporels sont meilleurs en performances absolues, mais aussi relativement à leur coût.