

Contextualisation de la sélection automatique de caméras : diffusion d'un match de basketball

Florent LEFÈVRE, Vincent BOMBARDIER, Patrick CHARPENTIER, Nicolas KROMMENACKER

Université de Lorraine, CRAN CNRS UMR 7039,
F-54506 Vandoeuvre-les-Nancy, France
Prenom.Nom@univ-lorraine.fr

Résumé – Cette étude propose une méthodologie de sélection automatique de caméras se basant sur la modélisation du contexte d'application. L'intégration de connaissances sur l'évènement permet d'identifier les sources d'intérêts dans la scène. Cette méthodologie est appliquée dans le cadre de la diffusion d'un match de basket. La sélection automatique de la caméra est basée sur la localisation de l'action d'intérêt et la détection des lancers-francs dans les flux vidéos.

Abstract – This study proposes a methodology for automatic camera selection based on the application context modeling. The integration of knowledge about the event makes it possible to identify the sources of interest in the scene. This methodology is applied in the context of the broadcasting of a basketball match. The automatic camera selection is based on the location of the action of interest and the detection of free throws in the video streams.

1 Introduction

La diffusion d'un évènement sportif d'échelle nationale est extrêmement coûteuse. De nombreuses caméras sont nécessaires pour couvrir l'ensemble du terrain et plusieurs personnes sont requises pour manipuler ces caméras. De plus, un régisseur est requis afin de sélectionner les images à diffuser parmi toutes les vues existantes. Les grands clubs sportifs peuvent se permettre l'emploi d'une équipe de production, contrairement aux petits clubs qui n'ont pas les moyens de s'offrir un tel dispositif.

Afin de permettre la réduction des coûts de diffusion, l'utilisation de systèmes de montages automatiques a été proposée. L'objectif est de proposer un flux vidéo monté à partir d'images provenant de plusieurs caméras installées autour de la scène. Un système de montage automatique est, en général, composé de trois étapes [1]. La planification consiste à extraire des informations sur la scène. Afin d'orienter les caméras vers la source d'intérêt, une étape de contrôle est nécessaire. Enfin, l'étape de sélection, consiste à choisir à partir des informations de la scène, une vue contenant la source d'intérêt. Le choix des caractéristiques à extraire est donc un élément important d'un système de montage automatique.

Dans cet article, la mise en place d'un système de sélection automatique de caméras pour la diffusion d'un match de basketball est étudiée. Différentes méthodes ont été proposées dans la littérature pour la réalisation d'un montage automatique de match de basketball [2, 3, 4]. Cependant, ces systèmes dépendent fortement de l'installation des caméras (position, angle de vue, nombre) rendant leurs utilisations difficiles dans d'autres contextes. De plus, la plupart de ces méthodes se concentrent sur la production de contenu, sans prendre en

compte les préférences des utilisateurs. La personnalisation du flux monté n'est que très rarement abordée et est, lorsqu'elle est proposée, souvent limitée à quelques choix.

Nous proposons de modéliser les connaissances sur le contexte d'application dans le but de les exploiter pour la mise en œuvre d'un système de montage automatique. Nous nous intéressons dans cet article uniquement aux étapes de planification et de sélection d'un système de montage automatique. Nous formalisons les connaissances sur les matchs de basketball à l'aide de la méthode NIAM/ORM, afin de trouver les différentes sources d'intérêt que souhaitent voir les spectateurs lors d'une diffusion. L'intégration de connaissances nous permet de proposer, une méthode de localisation de l'action dans le match permettant la sélection de caméras d'intérêt en direct. Une méthode de détection de lancer-franc est également présentée afin de proposer un premier levier de personnalisation du flux vidéo.

2 Contextualisation du montage

Lors de la mise en place d'un système de montage automatique, il est primordial de comprendre ce que veut voir le spectateur. En effet, l'identification des éléments importants pour les spectateurs permet de mettre en place les méthodes pour trouver leurs occurrences. Pour ce faire, il est nécessaire d'acquérir des informations sur le contexte à partir de différentes sources telles que des textes, des bases de données, ou des expertises humaines. L'intégration de connaissances est donc nécessaire pour proposer une méthodologie générique de montage automatique, afin de déterminer les sources d'intérêt dans

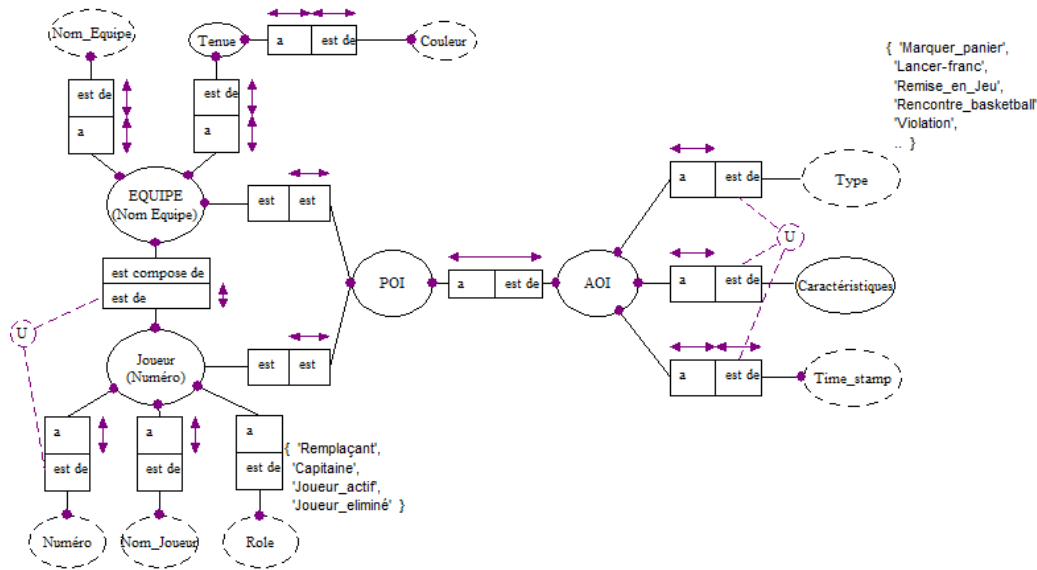


FIGURE 1 – Modélisation des sources d'intérêt d'un match de basketball

la scène. L'exploitation des connaissances sur le contexte permet également d'aider à mettre en place des méthodes pour la détection de ces éléments. Dans cet article, nous nous basons sur le règlement proposé par la Fédération International de Basketball Amateur (FIBA) [5] qui rassemble l'ensemble des connaissances sur le déroulement d'un match de basketball. Prenons l'exemple de la phrase suivante, définissant un match de basketball : "Une rencontre de Basketball se dispute entre deux équipes de 5 joueurs chacune. L'objectif de chaque équipe est de marquer dans le panier de l'adversaire et d'empêcher l'autre équipe de marquer." Nous pouvons extraire de cette phrase deux groupes distincts : les personnes et les actions. Une personne d'intérêt est une "équipe" composée de 5 "joueurs". Ces joueurs participent à l'action "rencontre de Basketball" qui est composée entre autre de l'action "marquer dans le panier" et "empêcher l'autre équipe de marquer".

Les informations que nous avons extraites du règlement officiel de la FIBA ont été modélisés par la méthode NIAM/ORM [6] (figure 1). Le principal intérêt de l'utilisation de cette méthode réside dans sa capacité à permettre l'extraction des connaissances à partir d'un texte en langage naturel et de les exprimer sous forme de relations simples ainsi que par sa facilité de formalisation à l'aide de notations graphiques compréhensibles pour un non-expert [7].

La modélisation des connaissances met en évidence deux concepts : les personnes d'intérêt (POI) et les actions d'intérêt (AOI). Nous nous intéressons, dans la section 3, à extraire l'AOI principale du match, appelée jeu notable, permettant une sélection de caméra en temps-réel. Afin d'offrir une diffusion personnalisée, il est nécessaire d'extraire des informations sur les différentes AOI et POI. Nous présentons, dans la section 4, une méthode de détection d'une des AOI, "les lancers-francs", se basant sur les connaissances extraites du règlement. Les informations extraites permettent alors de proposer un flux vidéo contenant uniquement les lancers-francs d'un match.

3 Localisation du jeu notable

Dans le cas de la diffusion en direct d'une rencontre de basketball, l'AOI principale est identifiée comme le jeu en lui-même. La caractéristique essentielle à obtenir est sa localisation. Cette information peut être obtenue de trois façons différentes : l'analyse de la direction des regards des spectateurs [2], la localisation de la balle [8] ou la localisation des joueurs [4]. L'analyse de l'orientation du regard des spectateurs permet de localiser facilement le point d'intérêt des spectateurs lors d'un match. Cependant, cela requiert l'utilisation de caméras supplémentaires, orientées vers les spectateurs et un temps de calcul important. La balle est au centre de l'attention des joueurs et sa localisation est représentative de l'action dans un match. Cependant, ses déplacements rapides, sa taille et les occlusions fréquentes rendent sa détection difficile avec un faible nombre de caméras. Les joueurs sont plus facilement détectables du fait d'une taille plus importante. Il est alors possible de détecter le jeu d'intérêt en analysant leurs déplacements. Cependant, les méthodes existantes requièrent l'utilisation d'un grand nombre de caméras, augmentant le coût matériel et calculatoire.

Afin de suivre l'action, nous proposons une méthode basée sur l'analyse du centre de gravité des joueurs. En effet, les spectateurs voudront voir les attaques et les défenses des différentes équipes. Du fait que cinq joueurs d'une équipe soient en attaque et que les cinq autres joueurs de l'autre soient en défense [5], la position de l'action d'intérêt est proche de la moyenne des positions des joueurs. La détection du jeu notable repose donc sur le déplacement de l'ensemble des POI. Le détail de la méthode utilisée et de son évaluation est disponible dans l'article [9]. Dans notre cas d'application (figure 2), une caméra azimutale permet de suivre le déplacement des joueurs. L'analyse du mouvement des joueurs permet de sélectionner une des trois caméras latérales filmant l'action. Pour ce faire, nous avons dé-

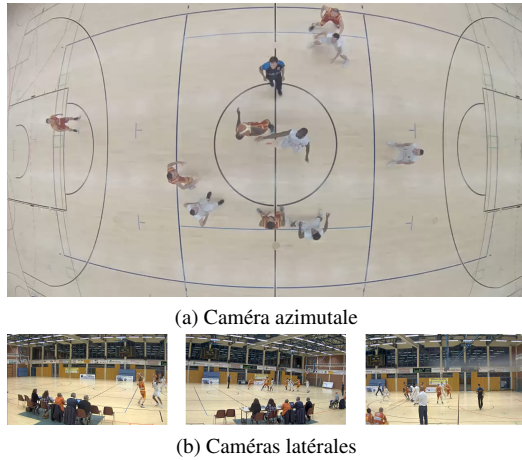


FIGURE 2 – Installation des caméras

fini manuellement trois zones dans l'image azimutale correspondant aux zones filmées par les caméras latérales. Lorsque le centre de gravité change de zone, la caméra correspondante est sélectionnée.

Nous pouvons voir dans la figure 3 deux montages différents correspondant aux 100 premières secondes du premier quart-temps d'un match de basket-ball. La courbe rouge correspond à un montage où le monteur sélectionne manuellement les caméras, en fonction des phases de jeu. La courbe verte correspond au montage automatique résultant de l'application de notre méthode.

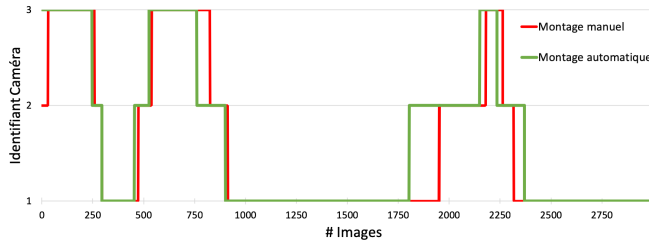


FIGURE 3 – Sélection de caméra au cours du temps

Les flux vidéos générés sont similaires à 86,2%. Les différences se situent principalement dans les transitions et proviennent du fait que le monteur aura tendance à anticiper ou retarder le changement de caméra. Sur un autre jeu de test, correspondant à une rencontre captée avec 4 caméras, les flux vidéos sont similaires à 76,3 %. Dans les deux cas, le changement de caméra par notre méthode est effectuée en moyenne 23 images (soit 0.76s) avant celle effectuée manuellement. Le temps de traitement par image est de 7 à 8 ms (130 fps), ce qui est compatible avec une diffusion en direct.

4 Détection de lancer-franc

Comme le montre le modèle NIAM/ORM, une action d'intérêt est définie par un type, une durée et un certain nombre de caractéristiques qui lui sont propres. L'étude du règlement offi-

ciel nous permet d'obtenir des informations sur des actions particulières. Par exemple l'AOI "lancer-franc", définie dans l'article 43 du règlement [5], comporte de nombreuses règles permettant de la reconnaître. L'AOI "lancer franc" possède ainsi plusieurs caractéristiques précises sur la position que doivent avoir les joueurs pendant toute sa durée. La figure 4 résume la position des joueurs lors d'un lancer-franc :

- Le tireur de lancer-franc ($A3$) doit se placer derrière la ligne de lancer franc, à l'intérieur du demi-cercle.
- Les joueurs occupant les places de rebond ($A1, A2, B1, B2, B3$) doivent prendre des positions alternées dans ces places.
- Les autres joueurs ($A4, A5, B4, B5$) doivent rester derrière la ligne de lancer-franc prolongée et derrière la ligne de panier à 3 points jusqu'à la fin du lancer-franc.

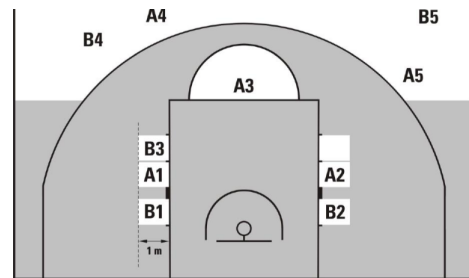


FIGURE 4 – Positions des joueurs durant les lancers-franc [5]

Ces informations sont utilisées pour mettre en place une méthode de détection de lancers-franc. En tenant compte les positions attendues lors d'un lancer-franc, nous pouvons définir manuellement, les zones correspondantes (figure 5). Du fait que nous utilisons une caméra azimutale de type fish-eye, il est nécessaire d'adapter ces zones afin de mettre en évidence les situations de lancers-franc. La figure 5 présente un lancer-franc détecté. Les joueurs participant à l'action sont dans les positions de lancer-franc tandis que les autres joueurs se trouvent derrière la ligne de lancer-franc. Les règles énoncées de la FIBA,



FIGURE 5 – Détection d'un lancer franc

stipulent que cinq joueurs se trouvent dans les zones de rebonds. Cependant, du fait de la déformation de l'image, nous ne pouvons garantir que les 5 joueurs soient détectés. Certains joueurs peuvent se trouver en dehors de l'image, d'autres joueurs peuvent être masqués par les joueurs les plus proches de la zone de lancer-franc (les joueurs $A2$ et $B3$ dans la figure 4).

Afin d'éviter des fausses détections lorsque les joueurs traversent les zones de lancer-franc, un délai peut être mise en place dans la détection. En effet, nous pouvons considérer la détection d'un lancer-franc, si les joueurs sont dans les bonnes positions pendant un certain nombre d'images successives. Le tableau 1 présente les résultats obtenus, en termes de précision, de rappel et de F-mesure pour différentes vidéos, en prenant en compte différents intervalles d'images. Les extraits 1 et 2 sont des vidéos comprenant 1 ou 2 lancers-francs et ont une durée de 1 minute 30 (2250 images) pour le premier et 6 minutes (9000 img.) pour le second. Les deux autres vidéos sont respectivement le premier quart-temps (15m18 - 22950 img.) et le second quart-temps (16m42 - 25050 img.) d'un match.

TABLE 1 – Impact du délai sur la détection des lancers-francs

Vidéos	Délai	Précision	Rappel	F-mesure
Extrait 1	1	90,82	75,20	82,26
	5	97,02	65,20	77,99
	10	100	54	70,13
Extrait 2	1	99,53	73,70	84,69
	5	100	61,25	75,96
	10	100	48,44	62,26
1 ^{er} QT	1	81,87	75,81	78,72
	5	88,54	66,97	76,26
	10	92,58	59,75	72,63
2 ^{eme} QT	1	84,78	81,12	82,91
	5	92,30	70,22	79,76
	10	95,74	62,25	75,44

Nous pouvons remarquer que plus le délai accordé pour valider un lancer-franc est grand, plus la précision augmente. En effet, son augmentation permet de ne pas prendre en compte les passages des joueurs dans les zones de lancers-francs. Cependant, cela implique qu'un vrai lancer-franc sera détecté avec un petit délai, ce qui se vérifie avec une diminution du rappel. L'étude de la f-mesure montre qu'un délai faible améliore globalement la détection de lancer-franc. Cependant, en terme de rendu, il est sans doute préférable de ne montrer que les vrais lancer francs quitte à en manquer le début (mise en place). La date de début du lancer franc pourrait de toute façon être avancée arbitrairement d'un temps correspondant au délai.

5 Conclusion

Dans cet article, nous proposons une méthodologie de montage automatique basée sur le contexte. L'étude du règlement de la FIBA nous a permis de comprendre comment se déroulait un match de basketball, d'en extraire et d'en modéliser les principales connaissances. Cette modélisation fait apparaître un grand nombre de sources d'intérêt séparées en deux catégories : les personnes d'intérêts et les actions d'intérêts. Pour proposer une diffusion en direct, nous avons proposé une nouvelle méthode de détection de l'AOI "jeu notable", fonctionnant en

temps-réel et simple à mettre en place. En vue de proposer la personnalisation du flux monté, il est nécessaire d'extraire plus de connaissances sur la scène. La détection des AOI et des caractéristiques des POI permet de proposer la personnalisation des flux vidéos montés. Nous avons proposé une méthode pour la détection de l'AOI "Lancer-franc", permettant de fournir aux spectateurs un résumé des lancers-francs d'un match. Les connaissances sur le contexte nous ont permis de mettre en place une méthode précise et fonctionnant en temps réel. L'intégration des connaissances peut permettre la reconnaissance d'autres AOI/POI facilement. Il est par exemple possible de distinguer une équipe par la couleur des maillots ; les joueurs sont eux identifiables par leur numéro qui est unique.

Enfin, notre méthodologie est applicable à d'autres types d'évènements tels que des conférences, des rencontres sportives ou des conseils municipaux [10].

Références

- [1] J. Chen and P. Carr. Autonomous Camera Systems : A Survey. In *Workshop on Intelligent Cinematography and Editing*, pages 18–22, 2014.
- [2] S. Daigo and S. Ozawa. Automatic pan control system for broadcasting ball games based on audience's face direction. In *21st ACM Int. Conf. on Multimedia*, pages 444–447, 2004.
- [3] F. Daniyal, M. Taj, and A. Cavallaro. Content and task-based view selection from multiple video streams. *Multimedia Tools and Applications*, 46(2-3) :235–258, 2010.
- [4] P. Carr, M. Mistry, and I. Matthews. Hybrid Robotic/Virtual Pan-tilt-zom Cameras for Autonomous Event Recording. In *21st ACM Int. Conf. on Multimedia*, pages 193–202, 2013.
- [5] FIBA. REGLEMENT OFFICIEL DE BASKETBALL, September 2018.
- [6] Terry A. Halpin. ORM/NIAM Object-Role Modeling. In *Handbook on Architectures of Information Systems*, pages 81–101. 1998.
- [7] B. Almecija. *Apports des connaissances métiers pour la conception d'un système multi-capteurs de tri qualité Billon*. PhD thesis, Université de Lorraine, Nancy, 2013.
- [8] Y. Arika, S. Kubota, and M. Kumano. Automatic Production System of Soccer Sports Video by Digital Camera Work Based on Situation Recognition. In *Eighth IEEE Int. Symp. on Multimedia*, pages 851–860, 2006.
- [9] F. Lefevre, V. Bombardier, P. Charpentier, N. Krommenacker, and B. Petat. Automatic camera selection in the context of basketball game. In *8th Int. Conf. on Image and Signal Processing*, 2018.
- [10] F. Lefevre, V. Bombardier, P. Charpentier, N. Krommenacker, and B. Petat. Automatic video stream selection method by on-air microphone detection. In *Int. Conf. on Pattern Recognition and Artificial Intelligence*, 2018.