

Synthèse de lectures des recherches du groupe FXPAL sur l'analyse et l'enregistrement de réunions

Seminaire SH

Étudiant L. Genasci
Université de Fribourg
Faculté des Sciences Informatiques
luca.genasci@unifr.ch

3 mars 2005

Résumé

Cet article a été écrit pour le séminaire du *Master* à l'Université de Fribourg, pendant le semestre d'hiver. Le groupe de recherche **DIVA**¹ de l'Université de Fribourg a collaboré au séminaire.

Cet article explique trois projets développés par le laboratoire de **FXPAL**.

Le premier projet donne une possible solution pour suivre, enregistrer et explorer les réunions.

Tandis que le deuxième parle d'une application visant à indexer une vidéo.

Enfin le dernier projet est une application qui permet d'accéder à des vidéos, le but est celui d'avoir plusieurs informations sur un certain produit.

Introduction

Dans cet article je parle de trois projets qui ont été développés dans le laboratoire du **FXPAL**. Ce laboratoire² se trouve à Palo Alto et il est partagé par deux entreprises : **Xerox** et

Fuji. **FXPAL** s'occupe de la recherche sur les technologies multimédia.

La **première partie** de ce document fait référence à l'article [1], qui propose un type de salle de réunions. Il y a trois moments importants pendant une réunion : pouvoir la suivre et l'enregistrer dans le même temps, ensuite lorsque la réunion a été enregistrée, avoir l'opportunité d'accéder à celle-ci.

La **deuxième partie** de ce document fait référence à l'article [3], qui propose un système préparé pour indexer une vidéo : la *Vidéo Manga*. Le principe du système est celui d'avoir un index fait par des images, tout comme une *bande dessinée*.

La **dernière partie** se réfère à l'article [2], qui décrit une application appelée *Mini Media Surfer*. Ce projet permet d'accéder à des vidéos explicatives sur un certain produit.

1 Meeting Room

1.1 Salle de réunions

La salle des réunions développée par **FXPAL** est utilisée pour plusieurs types de réunions : par exemple pour des réunions du personnel, pour des discussions et révisions de

¹<http://diuf.unifr.ch/diva>

²FXPAL = Fuji Xerox Palo Alto Laboratory

projets, pour des vidéoconférences, et pour des présentations.

1.1.1 Structure de la salle

Dans la salle des réunions du **FXPAL**, il y a divers médias (on peut les voir dans la figure 1). Tous les instruments ne sont pas visibles dans la figure, il y a en plus : les microphones qui se trouvent sur le plafond de la salle et deux caméras vidéos.

Les composantes qui ne sont pas nécessaires pour les conférences sont placées à l'extérieur de la salle : cela permet d'éliminer le problème des bruits³. Le projecteur aussi est mis dans une salle à côté pour éviter que son bruit dérange trop l'orateur et les participants.



FIG. 1 – Salle de réunions du **FXPAL**.

La caméra vidéo pour la conférence à distance⁴ est mise à côté du tableau pour la projection.

Le tableau pour la projection⁵, est le média le plus important : sur ce tableau est projetée la présentation de l'orateur.

Tantis que le tableau blanc⁶ est utilisé pour prendre des notes pendant les réunions.

Ensuite il y a une caméra vidéo pour les documents imprimés sur papier ou pour des notes

³par exemple le bruit des ordinateurs

⁴voir la figure 1, étiquette : *Videoconference camera*

⁵voir la figure 1, étiquette : *Rear projector*

⁶voir la figure 1, étiquette : *Whiteboard*

sur feuille prises à la main.

Les trois caméras vidéo⁷ de la salle, peuvent être bougées : normalement une caméra vidéo est pointée sur le tableau blanc, une autre est pointée sur l'orateur, et la dernière fait une panoramique de la salle. Lorsqu'un veut faire une prise de vue particulière, il y a une salle de contrôle à côté de la salle des réunions que nous propose ce service (on peut l'entrevoir dans la figure 2, dans le fond de la salle où il y a une fenêtre).

Enfin les *TabletPC*⁸ sont utilisés par chaque participant, pour suivre la réunion dans la salle.

En dernier lieu, dans la salle, il y a la place destinée à l'orateur ; elle est appelée *Podium* (on peut voir dans la figure 2).

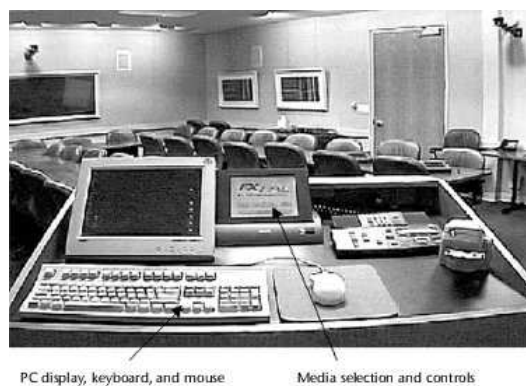


FIG. 2 – Le *Podium* - Place de l'orateur.

Sur le *Podium*, il y a plusieurs outils pour gérer la présentation : un ordinateur dans le quel ont été mises les présentations, une interface de contrôle⁹ pour sélectionner le média, la caméra vidéo pour les documents¹⁰ sur papier, et une place pour connecter le laptop personnel.

⁷voir la figure 1, étiquette : *Room camera (1 of 3)*

⁸voir la figure 1, étiquette : *Wireless pen computers*

⁹voir la figure 2, étiquette : *Media selection and controls*

¹⁰voir la figure 1, étiquette : *Document camera*

1.2 Suivre une conférence

Lorsque l'orateur commence sa présentation les participants peuvent la suivre en *real-time* sur sont *TabletPC*. Pendant la conférence ils ont aussi la possibilité de prendre des notes sur le matériel proposé. Pour rendre le travail de prise des notes encore plus efficace, aux participants est donné la possibilité de sélectionner et prendre un échantillon représentatif d'une caméra vidéo. À fin de pouvoir profiter de cet outil le participant doit posséder un'application appelée *NoteLook client*¹¹.

1.2.1 NoteLook client

Cette application a été développée pour suivre et prendre des notes pendant la réunion. On peut voir l'application dans la figure 3.

NoteLook client utilise le modèle YCAGWYS (*You Can Always Get What You See*). Le sigle veut nous expliquer que dans l'application *NoteLook client* on retrouve exactement ce qu'on voit dans la salle au même moment.

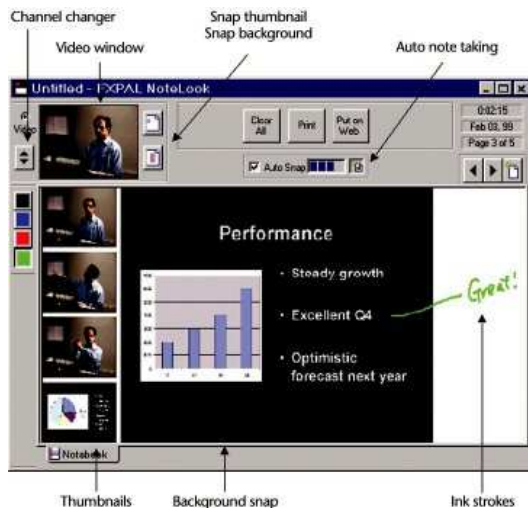


FIG. 3 – Application *NoteLookClient*.

¹¹développe par FXPAL

1.2.2 Comment utiliser l'application

L'application se présente comme dans la figure 3. Si l'on regarde l'application en haut, à gauche, il y a un *radio-button* qui permet de se joindre à la réunion. Tout de suite, en bas, on trouve un bouton qui change le canal du média¹². À droite de ces boutons, il y a une fenêtre où le média sélectionnée est reproduit. Le bouton qui se trouve à droite est utilisé pour prendre un échantillon représentatif du média selectionné. Dans ce dernier cas, une fois l'action faite, l'échantillon sera empilé à gauche de l'application¹³. Avec le stylo du *TabletPC* on peut prendre des notes numériques sur tous les échantillons prises lorsque ceux-ci sont affichés sur le *Background snap*. Dans la partie supérieure de l'application il y a un bouton qu'une fois appuyé permet de prendre des échantillons automatiquement¹⁴. Ce dernier a comme but de faciliter la prise des échantillons : en pratique, lorsque l'orateur change le transparent ou le média, le système empile un échantillon représentatif dans le *thumbnails*. Lorsqu'un participant prend un échantillon, ou des notes, le temps de l'action sera enregistré par le système.

Toutes les notes, les petites images¹⁵ et les transparents seront utilisés pour indexer la conférence, et ils sont synchronisés aux médias.

1.3 Navigation de la conférence

Une fois que la conférence est terminée, l'application *NoteLook client* peut créer le code HTML qui permet de la publier sur le Web.

La réunion est indexée par les notes et les échantillons pris par les participants. Toutes les données enregistrées sont disponibles sur le Web comme on peut le voir dans la figure 4.

¹²voir la figure 3, étiquette : *Channel changer*

¹³voir la figure 3, étiquette : *Thumbnails*

¹⁴voir la figure 3, étiquette : *Auto note taking*

¹⁵échantillons



FIG. 4 – Application Metadata media Player et les pages HTML.

Sur les pages HTML les photos et les notes prises pendant la conférence sont liées au vidéo : lorsqu'on clique sur un lien, on ouvre une application appelée Metadata Media Player¹⁶, et on voit la vidéo dans l'instant où l'on a pris des notes ou un échantillon représentatif.

1.4 Architecture du Système

Pour faire fonctionner tout ce que j'ai présenté et décrit jusqu'ici, on doit posséder un système compliqué qui traite toutes les données relevées pendant une réunion.

Un schéma représentatif du système est reproduit dans la figure 5.

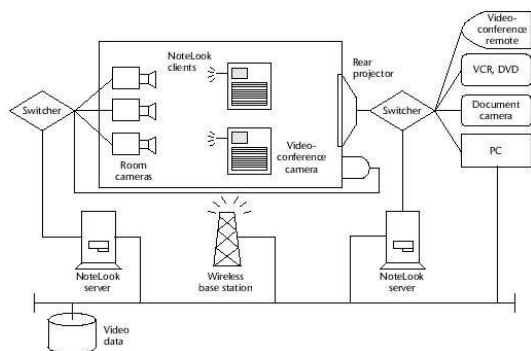


FIG. 5 – Architecture du système.

¹⁶voir la figure 4

Le carré central de la figure 5 est la salle de réunions. Les outils tels que les NoteLook servers, la base de données et la station wireless sont à l'extérieur. Deux raisons expliquent ce déplacement : tout d'abord ces instruments font beaucoup de bruit et ils dérangent l'enregistrement du son pendant la conférence, en outre cela permet d'avoir plus de liberté de mouvement dans la salle.

La station wireless sert à communiquer les événement plus importants aux participants.

Les switchers contrôlent les flux des médias qui sont présents dans la salle de la conférence.

Les NoteLook servers ont la responsabilité de prendre et traiter les flux des médias présents dans la salle, envoyer les information à l'application NoteLook client et enregistrer et synchroniser toutes les données à fin de pouvoir les récupérer plus tard.

Le système a des autres caractéristiques importantes. Une des celles-ci est de comprendre le média qui possède le meilleur signal pour enregistrer le flux. Une autre est celle de détecter le moment du changement d'un transparent ou de la sélection d'un nouveau média : lorsqu'un événement entre ceux-ci s'ajoute, le système envoie un échantillon représentatif à l'application NoteLook client. Le réseau wireless a seulement un flux de 11 Mb et pour éviter sa saturation on envoie seulement 0.5 frame par seconde.

1.5 Conclusions

Ce système permettant l'enregistrement ces conférences est très flexible car il fourni l'orateur plusieurs outils informatiques, le sujet peut donc utiliser n'importe quel support pour sa présentation (sur papiers, PowerPoint, ...). Il simplifie son travail aussi du point de vue des déplacements, l'orateur ne doit plus bouger par la salle pour sélectionner les médias dont il nécessite, mais ceux-ci seront tous disponibles dans le podium. Avec l'interface de contrôle sur

le *podium* il est très facile de sélectionner les médias. En revanche pour ce qui concerne les participants, eux ils peuvent prendre des notes sur tous les médias disponibles dans la salle, tout en ayant pleine liberté de choisir la place où s’asseoir.

Le projet présenté a été très bien conçu, mais il y a quand-même des améliorations possibles : par exemple on pourrait projeter une reconnaissance tout du text sur les transparents qui sont exposés que de la voix. Si la partie d’une phrase dite par quelqu’un se trouve dans le transparent projeté, on doit avoir la possibilité de lier la phrase au vidéo. On peut également ajouter dans l’application *NoteLook client* un bouton pour les questions : lorsqu’un participant clique sur ceci, il se prénote pour une question. Le système avertira l’orateur et au moment favorable il autorisera le participant à intervenir. Dans le même instant, une caméra vidéo sera prête pour se bouger. Le participant pourra intervenir, lorsque les microphones relèvent sa position et le système enregistre le temps. Introduire ce système on peut améliorer l’index de la conférence. Dans les page *HTML* on peut mettre un échantillon représentatif de la question qui a été posée.

2 Video Manga

La *Vidéo Manga* est un système utilisé pour indexer une vidéo. Le résultat final est comme celui d’une *bande dessinée*¹⁷, où les images les plus grandes représentent les plus importantes tandis que les plus petites les moins importantes. Dans la figure 6 on peut voir un index d’une vidéo.

En premier lieu on regroupe dans un *cluster* tous les échantillons similaires, ensuite dans celui-ci seront déterminés les segments vidéos. La règle qui nous indique comment décider quels segments sont les plus importants se base

¹⁷comme un *manga*



FIG. 6 – Application Mini Media Surfer.

sur la fréquence et la durée d’un segment. Si le segment est rare et la durée (dans le temps) est longue, alors le segment est plus important, par contre s’il est répété plusieurs fois ou la durée est courte il sera moins important. L’image représentative pour l’index est choisie au moyen du segment.

3 Mini Media Surfer

Pour comprendre mieux cette application, j’en donne un scénario : je suis dans un magasin et j’ai besoin d’avoir des informations plus précises sur un certain article. Je prends mon *PDA* et je me connecte a la base de données de l’entreprise qui produit l’article dont je m’intéresse. J’interroge la base de données pour qu’elle puisse m’offrir une liste possible des vidéos, je peut en choisir une et la consulter pour acquérir des informations sur l’article.

L’application *Mini Media Surfer* du **FX-PAL** a été développée pour fonctionner sur des systèmes embarqués (par exemple les *PDA*). On peut voir dans la figure 7 comme se présente l’application.

Lorsque l’application est démarrée on voit qu’on peut mettre un mot pour la recherche



FIG. 7 – Application Mini Media Surfer.

dans la base de données. En suite, on clique sur le bouton *search* et la base de données propose une liste des vidéos possibles. À gauche de chaque titre, il y a une barre qui indique le rapport d'un mot présent dans la vidéo : plus la barre est haute et plus le mot est présent dans la vidéo.

Une fois qu'on a sélectionné une voix, il devient rouge et apparaît en transparence un échantillon représentatif de la vidéo : on pourra ensuite passer à la reproduction de la vidéo. En bas, il se trouve une autre barre : cette-ci représente le segmentation de la vidéo. Dans un segment on voit une autre barre qui indique le rapport du mot présent cherché avant. Enfin on peut choisir le segment désiré et passer à la reproduction.

Références

- [1] Patrik Chiu, Ashutosh Kapuskar, Sarah Reitmeier, and Lynn Wilcox. Room with a read view : Meeting capture in a multimedia conference room. *In IEEE Multimedia*, Volume 7 Issue 4 :pp. 48–54, 2000.
- [2] Maryam Kamvar, Shingo Uchihashi, Lynn Wilcox, Sandeep Casi, and Surapong Lertsithichai. Browsing video segments on small displays. *CHI 2004 short paper*, April 2004.
- [3] Shingo Uchihashi, Jonathan Foote, Andreas Girgensohn, and John Boreczky. Video manga : generating semantically meaningful video summaries. *In Proceedings of*

the seventh ACM International Conference on Multimedia, Orlando, Florida, United States, pages pp. 383–392, 1999.