HAL
open science

Face2Graph: Base de données graphe et visualisation pour l'annotation d'archives vidéos

Adrien Dufraux, Benjamin Renoust, Shin'Ichi Satoh

To cite this version:

Adrien Dufraux, Benjamin Renoust, Shin'Ichi Satoh. Face2Graph: Base de données graphe et visualisation pour l'annotation d'archives vidéos. Extraction et Gestion des Connaissances, 2017, Grenoble, France. hal-01454977

HAL Id: hal-01454977

https://hal.science/hal-01454977v1

Submitted on 3 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Face2Graph: Base de données graphe et visualisation pour l'annotation d'archives vidéos

Adrien Dufraux* ** ***, Benjamin Renoust* **, Shin'Ichi Satoh*

*National Institute of Informatics, Tokyo, Japan
**JFLI CNRS UMI 3527, Tokyo, Japan
***INSA, Rouen, France
adrien.dufraux@gmail.com
{renoust, satoh}@nii.ac.jp,

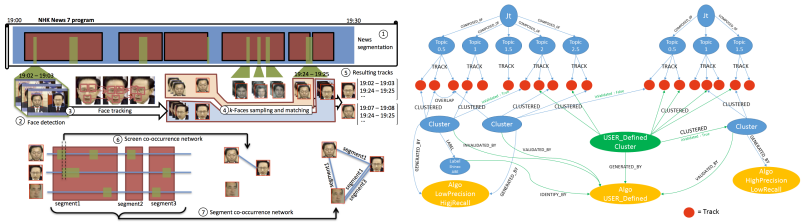
Résumé. Nous proposons dans ce travail d'utiliser la flexibilité des modèles de base de données graphe, et la représentation intuitive du réseau social afin de visuellement explorer, annoter, et vérifier des détections de visages dans une archive de 15 années de journaux télévisés.

1 Introduction

La réalisation d'analyses de haut niveau sur de grandes archives de documents (photos et vidéos), telles que des analyses médiatiques, sociales, juridiques ou politiques, demande souvent une phase d'annotation des documents (Renoust et al., 2016). Les technologies actuelles nous permettent de déployer à moindre frais des outils à grande échelle comme la détection et la reconnaissance de visages. Nous nous intéressons particulièrement à la reconnaissance de personnes dans une grande archive de vidéos de journaux télévisés (JTs). Bien que la reconnaissance de visages atteigne aujourd'hui de hauts niveaux de fiabilité, il est nécessaire de construire de grandes bases de vérité pour leur implémentation. Cela peut fonctionner pour les personnalités les plus connues, mais pour les cas les plus spécifiques, il est toujours nécessaire de passer par une étape d'annotation manuelle des visages. Nous proposons avec Face2Graph de faciliter cette tâche d'annotation via une information contextuelle sous la forme d'un réseau social de co-apparition dans les informations télévisées. Suivant le vieil adage *"dis-moi qui sont tes amis, je te dirais qui tu es"*, il est plus facile de trouver l'identité d'un individu au sein d'un groupe que de manière isolée. L'exploration du réseau de co-apparition avec Face2Graph propose une manière originale d'annoter les visages dans notre archive. La souplesse d'une modélisation en base de données orientée graphe permet de dériver et de modifier rapidement les liens entre visages et personnalités détectés dans notre archive de 15 années de JTs.

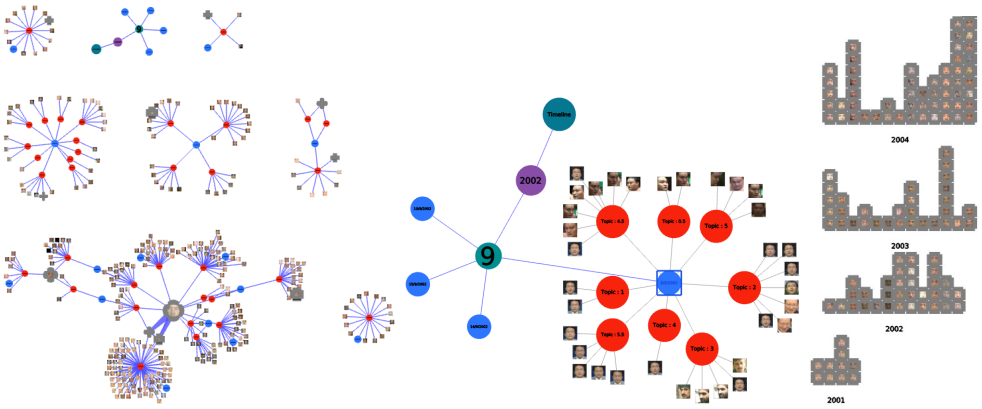
2 Présentation des données

Nous nous intéressons à l'annotation de groupements de visages dans une archive quotidienne de 15 années du journal de 19h (News7) de la chaîne public japonaise NHK (Katayama et al., 2005). L'archive est collectée depuis le 17 Mars 2001, soit plus de 5700 JTs et 7To de



(a) Analyse des vidéos (b) Architecture Neo4J

FIG. 1 – *Pré-traitement des JTs et structure de la base Neo4J, celle-ci peut intégrer ensemble différents clustering (modélisés en jaune).*



(a) Requête (b) Expansion d'une date (c) Histogramme

FIG. 2 – *(a) Vue d'ensemble de notre premier graphe entre le 28 août et le 15 septembre 2002. (b) L'utilisateur explore une date. (c) Un Cluster sur plusieurs années.*

vidéos. Notre objectif final est d'utiliser la détection automatique des individus dans cette archive pour effectuer des analyses sociologiques et médiatiques. Chaque JT est ainsi découpé en une série de sujets d'actualités (*Topic* dans la figure 1, hérité de (Ide *et al.*, 2004)).

Notre première étape consiste donc en une analyse préliminaire des données pour extraire et regrouper les visages dans cette archive (clustering), en 4 étapes (héritée de (Ngo *et al.*, 2013)) : **1.** Détection des visages avec Viola-Jones (Viola et Jones, 2004), **2.** Suivi des visages avec un traqueur de points Kanade-Lucas-Tomasi (Shi et Tomasi, 1994) formant des *tracks* (même visage détecté de manière consécutive), **3.** Échantillonnage des tracks, et description des visages dans un espace à 128 dimensions en utilisant OpenFace (Amos et al., 2016), **4.** Clustering des tracks avec GreedyRSC (Le et Satoh, 2011), un cluster de tracks correspondant à une seule et même personne. Le résultat final forme une collection de plus de 200k tracks, totalisant plus de 30M instances de visages. De plus, nous héritons d'une première annotation manuelle de 139 clusters de personnalités couvrant 16k tracks. Les imperfections du processus nous forcent à tenir compte de plusieurs situations : les tracks peuvent être dans un mauvais cluster, ou dans aucun cluster. Certains clusters peuvent être mal annotés, beaucoup ne le sont

pas. Une même personne peut être représentée par plusieurs clusters.

La richesse de ces données se situe dans les liens entre ces visages. A cause du paramétrage de nos algorithmes et des choix éditoriaux de la NHK, la détection de la co-apparition de personnes sur le même écran reste assez rare. Mais les co-apparitions dans un même sujet d'actualité sont très nombreuses et ont pour sémantique le fait que les personnes détectées sont concernées par un même sujet/thème. Un politicien apparaît souvent avec d'autres politiciens, influents, ou du même parti, un sportif avec d'autres sportifs de la même discipline *etc*.

Comme la quantité de données est importante, nous avons modélisé le tout dans une base de données graphe (Neo4J) afin de prendre en compte indépendamment plusieurs méthodes de regroupement. Ceci nous permet d'autoriser la correction des résultats précédents et la construction d'une base de vérité sans toutefois perdre l'information initiale de l'algorithme (pour potentiellement évaluer deux algorithmes différents). D'un point de vue interface, la présentation de toute la base de donnée présenterait une surcharge importante d'informations, et cette base nous permet tout particulièrement d'effectuer deux types de requêtes pour annoter un sous-corpus : par fenêtre temporelle ou par personne. Les liens entre clusters et/ou tracks sont simplement inférés depuis le découpage des JTs.

3 Visualisation et interaction

Les résultats d'une requête (dates ou personnes) sont présentés sous la forme d'un diagramme nœuds-liens représentant le réseau social des visages détectés, soit durant la période requise, soit en co-occurrence dans les sujets à partir des personnes requises. Un nœud correspond donc à un cluster de tracks, un lien à la co-occurrence dans les sujet, le dessin est ainsi celui d'un diagramme de force. Chaque cluster peut être exploré individuellement pour en vérifier le contenu et la temporalité. Beaucoup de tracks n'appartiennent à aucun cluster ou sont des répétitions des mêmes personnes. Afficher l'ensemble des liens rendrait la visualisation peu compréhensive. Nous proposons quelques regroupements et un encodage multiple afin d'alléger l'exploration : toutes les tracks isolées et non-identifiées sont regroupées autour d'un nœud de sujet (en rouge) correspondant au sujet du JT dans lequel ces visages ont été détectés. Les sujets sont connectés à un nœud de JT (en bleu) représenté par sa date.

L'exploration est centrée sur les clusters de tracks connus. Les clusters de tracks sont ainsi directement reliés entre eux lorsqu'ils co-occurrent au travers de sujets (l'épaisseur du lien représente le nombre de co-occurrences), ainsi qu'aux sujets avec lesquels ils partagent des tracks encore inconnues. Par souci de concision, nous avons "contracté" les nœuds des dates n'ayant aucune co-occurrence avec un cluster connu, de sorte à ce qu'elles ne soient représentées que part leur arbre temporel : date (bleu), mois (vert), année (violet), enraciné par un nœud intitulé "Timeline". Cette arborescence peut bien-sûr être éclatée à volonté par l'utilisateur souhaitant en explorer les tracks. Les nœuds de tracks individuelles sont carrés et présentent le (premier) visage détecté. Il en va de même pour les clusters annotés, carrés, dont le nom est affiché en dessous du nœud, ainsi que les clusters non-annotés qui eux sont représentés par un cercle. Les bordures des clusters validés sont vertes, rouges lorsqu'ils sont invalidés, et grises par défaut. Outre la validation, l'utilisateur peut par sélection fusionner tracks et/ou clusters ensemble (créant éventuellement un nouveau cluster), ou encore éditer le nom d'un cluster.

Chaque cluster est en réalité un méta-nœud, et l'exploration du contenu d'un cluster, se présente sous la forme d'une série d'histogrammes. Un histogramme par année, une colonne

par date de détection où sont accumulés des noeuds représentant chacun une track (donc le premier visage de chaque track). L'utilisateur peut extraire des tracks de celui-ci, ou même former un tout nouveau cluster à partir d'un sous-ensemble. Enfin l'utilisateur pourra décider d'annuler ses modifications, ou bien d'exporter sa base de vérité nouvellement formée. Le prototype a été réalisé avec Tulip (`tulip.labri.fr`) et Python pour l'interaction avec Neo4J.

4 Discussion et conclusion

Notre prototype présente une nouvelle façon d'explorer, de corriger, et d'annoter les résultats d'algorithmes de regroupement, grâce à une information contextuelle additionnelle (temporelle et co-occurrence). La nature de ces relations (poids et motifs d'associations temporelles) reste à intégrer dans ce système. Nous ne nous sommes intéressés pour le moment qu'aux visages, mais il est clair qu'encore plus d'informations contextuelles pourront être dérivées du contenu sémantique correspondant au sujet des JTs. Pour cela l'exploration d'une base de données graphe s'adapte parfaitement. Une extension idéale de ce travail sera directement implémentée en ligne dans un navigateur et mettra en lien direct détections et vidéos.

Références

Amos, B., B. Ludwiczuk, et M. Satyanarayanan (2016). Openface : A general-purpose face recognition library with mobile applications. Technical report, Technical report, Carnegie Mellon University-CS-16–118, Carnegie Mellon University School of Computer Science.

Ide *et al.*, I. (2004). Topic threading for structuring a large-scale news video archive. *Image and Video Retrieval 1*(1), 123–131.

Katayama, N., H. Mo, I. Ide, et S. Satoh (2005). Mining large-scale broadcast video archives towards inter-video structuring. *Advances in Multimedia Information Processing-PCM 2004 1*(1), 489–496.

Le, D. D. et S. Satoh (2011). Indexing faces in broadcast news video archives. In *2011 IEEE 11th ICDM Workshops*, pp. 519–526.

Ngo *et al.*, T. D. (2013). Face retrieval in large-scale news video datasets. *IEICE TRANSACTIONS on Information and Systems 96*(8), 1811–1825.

Renoust, B., T. Kobayashi, T. D. Ngo, D.-D. Le, et S. Satoh (2016). When face-tracking meets social networks : a story of politics in news videos. *Applied Network Science 1*(1), 4.

Shi, J. et C. Tomasi (1994). Good features to track. In *Pr. CVPR'94.*, pp. 593–600. IEEE.

Viola, P. et M. J. Jones (2004). Robust real-time face detection. *International journal of computer vision 57*(2), 137–154.

Summary

We propose in this work to take advantage of the flexibility of graph oriented databases and the intuitive visual representation of social networks to assist experts in exploring, annotating, and validating face detection and tracking in a 15 year archive of news programs.