

# Classification d'images télévisées

FARHAD - HEYBATI  
YOUCEF - KACER  
MARTIN - PROVOST

9 May 2016



# Table des matières

<b>Introduction</b>	<b>i</b>
<b>1 Images exploitées</b>	<b>1</b>
1.1 Corpus et labélisation . . . . .	1
1.2 Classification binaire . . . . .	3
<b>2 Extraction d'information</b>	<b>7</b>
2.1 Extraction de contours . . . . .	7
2.2 Extraction de teinte . . . . .	7
2.3 Histogramme orienté du gradient . . . . .	7



# Introduction

Ce document présente plusieurs méthodes d'extraction d'informations à partir d'images issues d'un débat télévisé. Cela afin d'en effectuer une classification binaire en 'gros plan' ou 'plan large'. Nous proposons dans un premier chapitre de présenter les images exploitées, et les deux classes qui nous intéressent. Dans une seconde partie, nous présentons les méthodes d'extraction d'informations utilisées. Puis dans une troisième partie, les résultats obtenus.



# Chapitre 1

## Images exploitées

### 1.1 Corpus et labélisation

Nous allons exploiter un total de 2351 images prises à partir de la vidéo d'un débat télévisé [Ess16]. Nous avons exploité le fichier de transcription au format .trs [Ess16] associé à cette vidéo, cela afin d'extraire les différentes classes, et attribuer à chaque image sa classe. En effet, le fichier de transcription fournit un total de 9 classes :

*M* : La présentatrice est seule à l'écran

*A* : La première intervenante est seule à l'écran

*B* : La seconde intervenante est seule à l'écran

*C* : Le premier intervenant est seul à l'écran

*D* : Le second intervenant est seul à l'écran

*ALL* : Les 5 personnes sont à l'écran

*MULTI* : Entre 2 et 4 personnes sont à l'écran

*INTRO* : Reportage d'introduction à l'écran

*CREDITS* : Générique d'émission à l'écran

Par ailleurs, le fichier de transcription donne à chaque intervalle de temps, ce qui est à l'écran à travers ces classes. Moyennant, une conversion du fichier au format xml, on peut obtenir le tableau suivant :

<b>classe</b>	<b>debut (s)</b>	<b>fin (s)</b>
<i>CREDITS</i>	0	11.36
<i>INTRO</i>	11.36	84.64
<i>M</i>	84.64	95.12
<i>ALL</i>	95.12	103.2
<i>A</i>	103.2	112.2
<i>M</i>	112.2	115.24
⋮	⋮	⋮
<i>M</i>	2334.12	2340
<i>ALL</i>	2340	2340.76
<i>CREDITS</i>	2340.76	2349.72

TABLE 1.1 – Table de correspondance classes/intervalle de temps



Les 2351 images étant prises à une seconde d'intervalle tout le long de la vidéo, on peut automatiquement labéliser celles-ci via la table de correspondance 1.1 vue en 2. Ci-après, nous présentons quelques images pour chacune des 9 classes :

## 1.2 Classification binaire

Par la suite, nous allons nous restreindre à seulement deux classes définies comme suit :

$G$  : 'Gros plan' (une seule personne est à l'écran)

$L$  : 'Plan large' (au moins deux personnes sont à l'écran)

Les classes  $G$  et  $L$  peuvent s'exprimer en fonction des 9 classes comme suit :

$G : M \mid A \mid B \mid C \mid D$

$L : ALL \mid MULTI$

Nous avons donc deux classes d'images  $G$  et  $L$  dont voici plusieurs exemples :

Par la suite, nous allons expliciter différentes manières d'extraire de l'information afin de pouvoir discriminer les images de classe  $G$ , des images de classe  $L$ .

FIGURE 1.1 – image de classe  $M$ FIGURE 1.2 – image de classe  $A$ FIGURE 1.3 – image de classe  $B$ FIGURE 1.4 – image de classe  $C$



FIGURE 1.5 – image de classe *D*



FIGURE 1.6 – image de classe *ALL*



FIGURE 1.7 – image de classe *MULTI*



FIGURE 1.8 – image de classe *INTRO*

FIGURE 1.9 – image de classe *CREDITS*FIGURE 1.10 – image de classe *G* (gros plan)FIGURE 1.11 – image de classe *L* (large plan)

## Chapitre 2

# Extraction d'information

### 2.1 Extraction de contours

### 2.2 Extraction de teinte

### 2.3 Histogramme orienté du gradient

Cette méthode consiste à extraire l'information local de contours. Sa première utilisation a consisté en la détection de piétons [DT05]. Pour une image donné, le vecteur descripteur des *HOG* est la concaténation d'histogrammes de l'amplitude du gradient (en fonction de l'orientation) pris sur un découpage de l'image :

On peut espérer que ces descripteurs s'adaptent bien à notre problème. En effet, les épaules des intervenants à gauche et à droite de l'image sont des contours discriminants pour la classe *G* (gros plan). D'autre part, les gros plans ont fond uniforme, ce qui donnera beaucoup de gradient nul, et donc un descripteur *sparse*

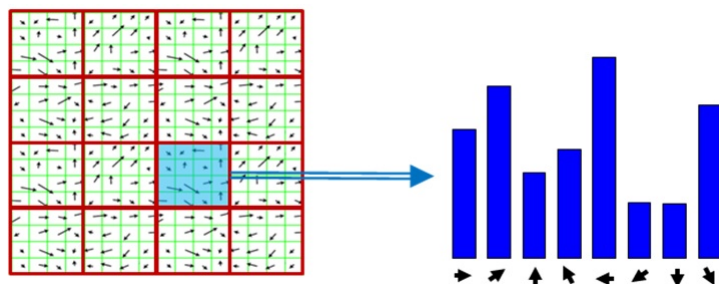
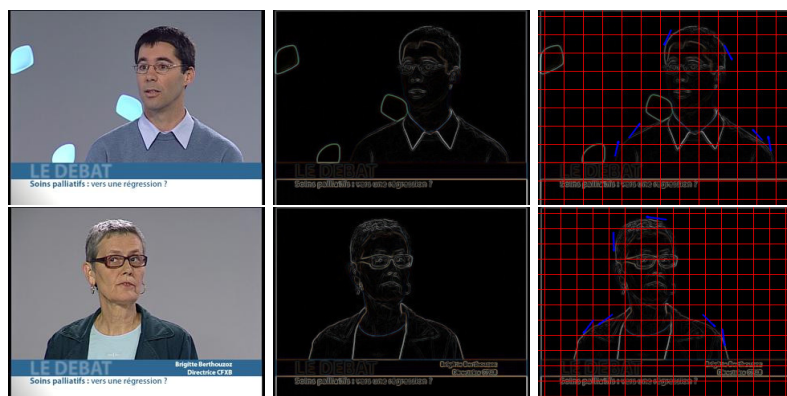


FIGURE 2.1 – construction d'un histogramme orienté du gradient [Lev13]

FIGURE 2.2 – histogramme orienté du gradient pour les gros plans (classe  $G$ )

# Liste des tableaux

1.1	Table de correspondance classes/intervalle de temps . . . . .	2
-----	---	---





# Table des figures

1.1	image de classe <i>M</i> . . . . .	4
1.2	image de classe <i>A</i> . . . . .	4
1.3	image de classe <i>B</i> . . . . .	4
1.4	image de classe <i>C</i> . . . . .	4
1.5	image de classe <i>D</i> . . . . .	5
1.6	image de classe <i>ALL</i> . . . . .	5
1.7	image de classe <i>MULTI</i> . . . . .	5
1.8	image de classe <i>INTRO</i> . . . . .	5
1.9	image de classe <i>CREDITS</i> . . . . .	6
1.10	image de classe <i>G</i> (gros plan) . . . . .	6
1.11	image de classe <i>L</i> (large plan) . . . . .	6
2.1	construction d'un histogramme orienté du gradient [Lev13] . . .	8
2.2	histogramme orienté du gradient pour les gros plans (classe <i>G</i> ) .	8



# Bibliographie

- [DT05] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. *cvpr*, jun 2005.
- [Ess16] Slim ESSID. Resources, 2016. [www.perso.telecom-paristech.fr/~essid/](http://www.perso.telecom-paristech.fr/~essid/).
- [Lev13] Gil Levi. A short introduction to descriptors, aug 2013. <https://gilscv-blog.com>.