

# Classificació automàtica d'imatges de comptadors de gas

Jose Ruano Zaragoza

**Resum**—Aquest projecte pretén estudiar la viabilitat d'un mètode per identificar els diferents tipus (marca i model) de comptadors de gas. Una representació d'imatges fonamentada en el model Bag of Words i un entorn enfocat al tractament d'imatges, com és OpenCV, ens possibilitarà l'opció de realitzar una anàlisi sobre l'efecte que produeix l'utilització de diversos paràmetres que permetin obtenir una representació diferent de característiques d'un mateix conjunt d'imatges, causant una repercussió directa als resultats finals de classificació. A més a més, s'han constatat millores significatives als resultats amb l'implementació d'una tècnica basada en la detecció de contorns, que permet reduir el soroll causat per elements no desitjats en les imatges.

**Paraules clau**— Bag of Words, Classificació d'imatges, Comptadors de gas, OpenCV.

**Abstract**—The aim of this project is to evaluate the viability of a method to identify different models (brand and type) of gas meters. A representation of images based on the model Bag of Words and an environment focused on image processing, such as OpenCV, will allow us the option to perform an analysis on the effect of the use of several parameters which allow obtaining a different feature representation of a same set of images, causing a direct impact to the final results of classification. Additionally, significant improvements in results have been noted with the implementation of a technique based on the detection of edges, which reduces noise for unwanted items in images.

**Index Terms**—Bag of Words, Gas meters, Image Classification, OpenCV.

---

## 1 INTRODUCCIÓ

COM reconeixem una imatge d'un cotxe, d'un cavall, o d'un arbre? Matemàticament, existeix una diferència entre una imatge que conté una carretera a una en que no hi és. Algorímicament, trobar aquesta diferència és l'objectiu de la classificació d'imatges.

Mentre la majoria dels mètodes de l'estat de l'art comparteixen certs components bàsics, no existeix tanta investigació a l'hora d'explorar sobre quins són els paràmetres òptims d'aquests components. La majoria d'aquests treballs tenen com a finalitat modificar o millorar un cert algorisme. Però sense haver obtingut uns resultats robusts d'aquest algorisme base, és més difícil poder certificar el valor que tenen aquestes millores.

Aquest treball es centra amb l'exploració i anàlisi del model *Bag of Words* [1], com a mètode de representació d'un diccionari d'imatges. Per mostrar els resultats, es farà servir el nostre mètode per resoldre el problema de la classificació automàtica d'imatges de comptadors de gas.

Un comptador de gas s'utilitza per mesurar el consum de gas que, tot i l'existència de comptadors digitals que

permeten obtenir la informació de manera digital, la seva difusió és lenta i, avui dia, a molts llocs encara es procedeix a fer la lectura de forma manual. Això implica un cost econòmic i de temps, que es pretén estalviar amb el simple processament d'una fotografia. El pas previ necessari, per a poder extreure la lectura d'una imatge d'un comptador, és el de reconèixer el seu tipus, és a dir, a quina marca i model pertany. Així doncs, l'objectiu principal d'aquest projecte és el d'explorar la viabilitat del model *Bag of Words* a l'hora de classificar un conjunt d'imatges de comptadors de gas, localitzant els paràmetres que permetin obtenir la representació de característiques més òptima.

L'organització d'aquest article és la següent: Un repàs a l'estat de l'art es pot trobar a la segona secció. La tercera secció detalla breument quin ha estat l'entorn utilitzat durant el treball. Durant la quarta secció s'explica quina ha estat la metodologia emprada. Una anàlisi i discussió detallada dels resultats es proporciona a la cinquena secció. Finalment, la sisena secció conté conclusions i les possibles línies a seguir a treballs futur.

## 2 ESTAT DE L'ART

La idea d'inspirar l'anàlisi i reconeixement d'imatges com adaptació de mètodes aplicats a la categorització de text a documents, com ara *Bag of Words* [1], no es nova.

- 
- E-mail de contacte: Jose.Ruano@e-campus.uab.cat
  - Menció realitzada: Computació
  - Treball tutoritzat per: Ernest Valveny Llobet (Computació)
  - Curs 2014/15

Zhu et al [2] va investigar la quantificació vectorial a petites zones d'interès quadrades d'una imatge, que va anomenar *keyblocks*, demostrant que aquestes característiques trobades produïen millors resultats que plantejaments anteriors basats amb descriptors purament semàntics – com ara el color o la textura.

Un descriptor acostuma a anar emparellat amb el seu detector, els quals localitzen les regions de la imatge amb més entropia o bé amb un cert criteri, com pot ser la detecció de cantonades Harris per Harris et al. [3], o el detector Dense [4], el qual “engraella” la imatge per a que totes les regions d'interès siguin igual de denses.

La investigació sobre els descriptors ha estat molt activa durant la darrera dècada, produint una millora substancial dels algorismes en velocitat de computació i robustesa davant variacions d'escala, rotació i soroll. Alguns exemples, inclouen SIFT per Lowe et al. [5], SURF per Bay et al. [6] i, més recentment, FREAK per Alahi et al. [7] que proposa un descriptor de punts d'interès inspirat en el sistema visual humà. Per la realització d'aquest projecte, com es mostra a la secció 4, el descriptor SURF ens ha proporcionat les característiques computacionals més favorables.

### 3 ENTORN

Durant el transcurs d'aquest projecte, s'ha fet servir `c++` com a llenguatge de programació i Visual Studio 2013 com a entorn de desenvolupament.

El codi inicial consistia en un mètode de cerca de text en imatges manuscrites. Aquest, es va modificar i adaptar lleugerament per poder-lo aplicar a imatges de comptadors de gas. Per l'implementació del nostre model, s'utilitza la llibreria *open source* OpenCV [8], la qual està centrada en el processament d'imatges en temps real i és integrable amb `c++`.

### 4 METODOLOGIA

Els principals passos que segueix el nostre mètode són, amb aquest ordre, els següents:

- Detecció i descripció de regions d'interès donat un conjunt d'imatges prèviament etiquetades.
- Assignació d'aquests descriptors d'imatges a un conjunt de clústers o centroides predeterminats, el que serà el nostre vocabulari, mitjançant l'algorisme *k-means* [9].
- Construcció d'una *Bag of Keypoints (BOK)* a partir del número de descriptors assignats a cada clúster (*matcher*).
- Entrenament del classificador multiclasse *Support Vector Machine (SVM)* [10] utilitzant la BOK com a vector de característiques.
- Repetició de les passes anteriors, experimen-

tant amb diferents tipus de parella detector/descriptor, variacions en els seus paràmetres, diferent tamany de vocabulari, o inclús amb la creació d'un petit programa per reduir el soroll no desitjat a les imatges proporcionades, a partir d'un model de detecció de contorns.

- Identificar i seleccionar quin ha estat el vocabulari més robust pel nostre conjunt d'imatges. Aquest serà el que, globalment, hagi suposat una millor precisió a la classificació.

Les següents subseccions i seccions, pretenen definir i discutir quines han estat les eleccions per a cada pas amb més detall.

#### 4.1 Extracció de característiques

El primer pas en el nostre procés és el de reduir la informació que tenim de cada classe d'imatges (array de píxels), de tal manera que, tot i haver-la reduït, mantinguem la seva representativitat.

Tot seguit es fa un repàs als detectors i descriptors que s'han utilitzat al treball.

##### 4.1.1 SIFT: Scale-Invariant Feature Transform

L'algorisme SIFT proposat per Lowe et al. [5], està enfocat en trobar quines són les característiques que descriuen una imatge invariants a escala i orientació, és a dir, característiques trobades a una imatge i a la mateixa imatge escalada, han de ser exactament les mateixes. Pot ser utilitzat com a detector o com a descriptor individualment, permetent possibles combinacions amb altres enfoc que només actuen com a detector (com per exemple el detector dense [4]) o amb descriptors com SURF [6].

**Detector:** Durant aquesta etapa, es tracta de detectar quines són les regions de la imatge invariants a canvi d'escala. Això s'aconsegueix generant un espai d'escala, que mitjançant un diferencia de gaussianes permeti seleccionar les regions més estables durant el procés.

**Descriptor:** Els *keypoints* es descriuen a partir d'un càlcul ponderat de l'orientació i magnitud dels gradients de totes les imatges en una regió propera al *keypoint*, com podem veure a la fig. 1 (a). Aquests descriptors es representen en forma de subregions 4x4 a un histograma (fig. 1 (b)). Així doncs, la mida final d'un descriptor SIFT és de 128 dimensions.

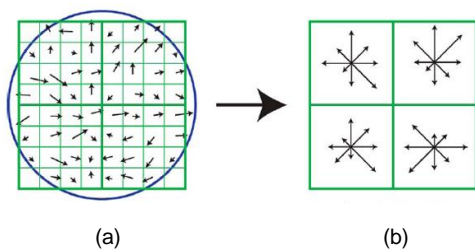


Fig. 1. Exemple de gradients i descriptors SIFT. (a) Es troben gradients a cada píxel (quadrat petits) i es ponderen en funció de la seva orientació i magnitud. (b) Descriptors resultants en forma de subregions 4x4. (figura extreta de [5])

#### 4.1.2 SURF: Speeded Up Robust Features

Surf per Bay et al [6], està reconegut com l'alternativa més eficient a SIFT. Proposa un tipus de detector basat en Hessians i una construcció de descriptors basada en la distribució. Al igual que succeeix amb el model SIFT, es pot fer servir el seu detector i descriptor de manera individual, permetent combinacions entre diferents mètodes d'extracció de característiques.

**Detector:** Es tracta de trobar regions característiques de les imatges a partir d'una aproximació de matrius Hessians.

**Descriptor:** Per aconseguir descriure els punts d'interès detectats, es fa servir un procediment semblant a la intensitat de gradient utilitzada al model SIFT. Aquesta, però, està basada en una distribució del *wavelet* de Haar prenent direccions horitzontals i verticals en comptes de la intensitat de gradients feta servir a SIFT. Es detalla el procés a la fig 2.

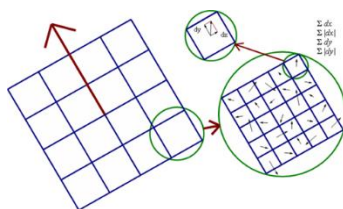


Fig. 2. Exemple de construcció de descriptors SURF. Es crea un quadrat perfecte al voltant del punt d'interès i es rota en funció de la seva orientació i es divideix en regions de tamany 4x4 (figura de l'esquerra). Aquesta regió es divideix en sub-regions iguals de tamany 5x5 a cada una de les quals es calculen les seves característiques. Obtenint un keypoint amb un tamany de 4 dimensions per a cada una de les sub-regions 4x4. (figura de la dreta). (figura extreta de [6]).

Concatenant la mida dels vectors-resposta del *wavelet* (4 dimensions) per cada regió 4x4, obtenim un descriptor SURF de 64 dimensions. Aquesta és una de les raons que fan el model SURF més ràpid enfront a les 128 dimensions ocupades pels descriptors SIFT.

#### 4.1.3 Dense feature detector

*Dense sampling*, és un detector de punts d'interès que

consisteix en dividir la imatge de manera que obtindrem una imatge amb regions igualment denses, resultant amb una mena d'engraellament de la imatge sobre 1 o varies escales (fig. 3).

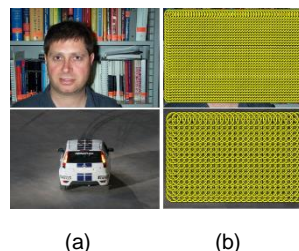


Fig. 3. Exemple d'engraellament dense. (a) Imatge original. (b) Imatge engrallada amb enfoc dense. (figura extreta de [4]).

## 4.2 Construcció d'un vocabulari

Idealment, un bon vocabulari ha de satisfer 2 criteris que, a priori, poden semblar contraposats. Per una part, ha d'incloure el menor nombre de característiques que continguin un elevat caràcter distintiu entre elles. Per altra banda, ha de ser suficientment gran com per contindre un descriptor, que permeti classificar cada possible *keypoint* trobat en un conjunt d'imatges amb la seva classe. És a dir, si dues característiques són visualment diferents no s'haurien de relacionar amb el mateix *keypoint* al vocabulari.

Per a la construcció del vocabulari s'utilitza l'algorisme *k-means*.

### 4.2.1 K-means

*K-means* Duda et al. [9] és un algorisme iteratiu de clustering que pren un número de punts  $k$  en un espai  $n$ -dimensional i els divideix entre regions delimitades per un número de clústers (centroïdes) més petit, maximitzant la distància entre posicions de clústers. A cada pas d'iteració, els punts es reassignen al clúster que tenen més a prop.

## 4.3 Bag of Keypoints

Anomenem *Bag of Keypoints* o *Bag of Features* [1] per analogia al mètode *Bag of Words* el qual refereix a la categorització de text a documents. En comptes de paraules (words) busquem punts o característiques d'interès coincidents.

Per a formar la nostra BOK farem servir un matcher que relacioni cada punt d'interès trobat a les imatges amb tots els *keypoints* existents al vocabulari i assignar-lo al més proper. Així doncs, la BOK portarà el compte del número de punts d'interès aparellats amb cada *keypoint* del vocabulari.

## 4.4 Classificació

Un cop realitzades les passes anteriors, estem preparats per començar el procés de categorització; predir les classes d'imatges no etiquetades. Aquest procés consta de

dos procediments ben diferenciats: training i testing.

Durant la fase d'entrenament es tracta d'enviar un conjunt d'imatges etiquetades al classificador, per tal de fer-lo aprendre (aprenentatge supervisat) a distingir entre diferents classes. En aquest projecte el classificador multi classe escollit ha estat l'SVM o *Support Vector Machines*, donat que és un dels classificadors més emprats a l'hora de resoldre problemes de tipus dimensionalment complexes, com és el nostre cas.

#### 4.4.1 SVM: Support Vector Machines

Els Support Vector Machines [10], són una eina d'aprenentatge computacional amb la capacitat de rebre certs vectors arbitraris com a entrada i tornar-los separats amb dos o més classes (fig. 4).

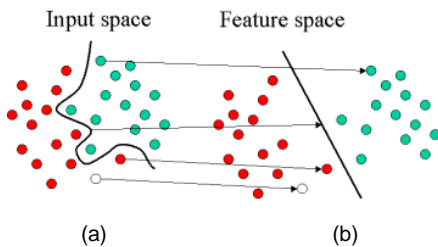


Fig. 4. SVM idea bàsica. (a) Durant l'espai inicial compost per vectors arbitraris, no es poden separar linealment. (b) Vectors separats linealment a l'espai kernel (figura extreta de [11]).

Utilitzant un espai addicional (kernel), es troba la distància lineal entre dos classes, que tenen com a frontera un hiperplà delimitat pel punt més proper de cada classe a aquest.

En el nostre cas, al ser un problema multi-dimensional, es tracta d'aplicar  $n$  cops el procés anterior SVM, sigui  $n$  el número de vectors (classes) d'entrada, i es compara l'espai kernel d'una classe amb una altra fins a successivament haver comparat totes.

#### 4.5 Reducció de soroll

Un dels principals escolts que es pot trobar un mètode com el nostre, tant a l'hora de crear el vocabulari, com a entrenar el classificador o categoritzar imatges, és el del soroll produït per objectes dels que no volem recavar informació i per tant distorsionen, en menys o més mesura, el resultat final (fig. 6).

Per tal d'apropar-nos al que seria una imatge més ideal (fig. 5) pel nostre model i que, a priori, doti d'una major efectivitat al procés, s'ha experimentat amb un mètode que s'aplica abans de l'execució del programa principal. Es tracta de retallar aquelles imatges en les que es pugui centrar més la forma del comptador, com les mostrades a la figura 6. Això es decideix en funció de contorns detectats depenent d'una certa intensitat llindar aplicada, aproximació dels contorns trobats amb capsos mitjançant l'algorisme Douglas-Pecker [12], i la posterior selecció d'aquestes capsos (*bounding boxes*) basada en les mesures d'un comptador.



Fig. 5. Exemple d'imatges existents al data set inicial més ideals per classificar.



Fig. 6. Diferent representació de quantitat de soroll a imatges del nostre data set.

Per tal de verificar el teòric benefici d'aquest mètode, es van proposar experiments de manera independent: amb el conjunt d'imatges originals i amb el mateix conjunt d'imatges retallades. A la secció de resultats es discuteix detalladament sobre els resultats obtinguts amb cada cas.

La fig. 7 mostra una imatge original, els contorns detectats per aquest mètode i el resultat final satisfactori un cop retallada.

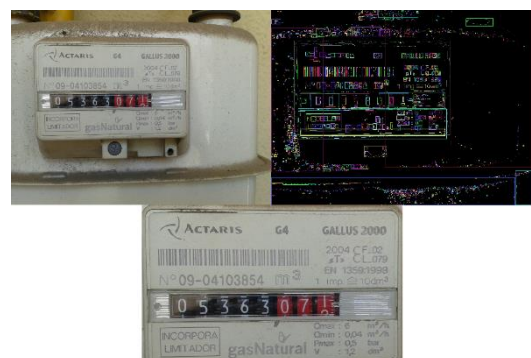


Fig. 7. Imatge original (adalt esquerra). Detecció contorns (adalt a la dreta) i resultat final satisfactori (imatge inferior).

La fig. 8 mostra un mal resultat amb el mateix procés.

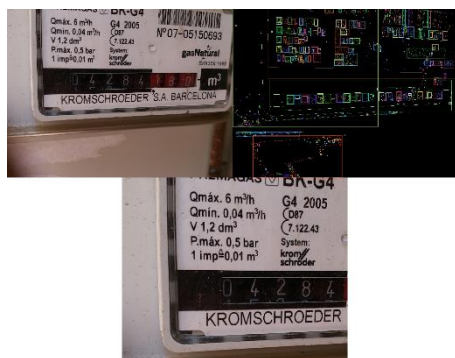


Fig. 8. Imatge original (adalt esquerra). detecció contorns (adalt a la dreta) i resultat erroni (imatge inferior).

## 5 RESULTATS

Tot seguit es presenten els resultats obtinguts als experiments. S'han dividit en tres apartats diferenciats que segueixen l'ordre cronològic en que es van realitzar. D'aquesta manera, el millor resultat a cada apartat ens serveix per continuar amb els experiments posteriors.

El primer apartat de resultats explora l'impacte produït per l'algorisme de reducció de soroll explicat a l'apartat 4.5. Es valoren els resultats produïts aplicant l'algorisme i sense aplicar-lo.

Prenent el millor resultat, el següent experiment proposa diferents tipus de combinacions entre detectors i descriptors.

El tercer i últim apartat, proposa experimentar amb diferent mida de vocabulari prenent com a configuració la millor obtinguda fins el moment.



Fig. 9. Alguns dels models de comptador de gas existents al nostre data set.

El nostre *data set* o conjunt d'imatges conté un total de 1183 imatges repartides en 20 classes. Cada classe reuneix els comptadors de gas del mateix tipus (marca i model). El contingut detallat de cada classe s'ensenya a la taula 1. La taula 2 detalla els millors resultats de precisió finals obtinguts a cada experiment. Uns quants exemples del nostre data set es mostren a la fig. 9.

TAULA 1  
IMATGES PER CLASSE

Classe	Total	Training	Test
53	51	41	10
55	34	27	7
58	62	50	12
62	37	30	7
68	60	48	12
69	43	34	9
70	64	51	13
71	59	47	12
72	65	52	13
73	56	45	11
75	66	53	13
81	59	47	12
82	52	42	10
83	63	51	12
86	70	56	14
88	64	51	13
89	75	60	15
90	68	55	13
91	56	45	11
94	79	63	16

Del total de cada classe, un 80% es dediquen a l'entrenament, i un 20% a testing.

S'ha d'afegir que el tamany del vocabulari ha estat de 999 elements o paraules durant tots els experiments, excepte a la secció 5.3 on específicament s'exploren resultats amb diferents valors.

TAULA 2  
COMPARATIVA AMB ELS MILLORS RESULTATS DE PRECISIÓ OBTINGUTS AMB CADA EXPERIMENT

Detector/Descriptor	Accuracy (%)
SURF/SURF	69.6
SIFT/SIFT	66.2
Dense/SURF	32
SURF/SURF(2000 paraules)	65

### 5.1 Reductor de soroll vs. Imatges originals

En primer lloc volem analitzar el rendiment de l'algorisme de reducció de soroll (secció 4.5) i decidir si ens val la pena utilitzar-ho o mantenir el data set original.

Durant aquesta etapa per a la construcció del vocabulari s'ha utilitzat l'extractor SURF, donat que, com s'explica a la secció 4.1.2, és el més ràpid en quant a temps de còmput.

Es mostren les matrius de confusió resultants amb el data set original i amb un data set al que se li ha aplicat l'algorisme de reducció de soroll a l'apartat A1 de l'apèndix.

A la taula 3, es pot observar com els resultats obtinguts amb el data set al que se li ha aplicat la reducció de soroll són un 3% més precisos (en general) que els obtinguts amb el conjunt d'imatges original.

Si ens hi fixem amb aquesta matriu de confusió, hi ha classes que localitzen tots o la gran majoria dels errors de classificació en una sola classe. Això és degut a que són comptadors de la mateixa marca però de diferent model i és possible que només es diferenciïn visualment amb el nom del model. La classe 58 per exemple erra 4 de les classificacions en la classe 55, o la classe 94 que confon 8 de les imatges i les classifica a la classe 89. A la fig. 10 s'exemplifica visualment aquesta situació.

Es van realitzar proves addicionals per veure si es mantenia la tendència agafant 5 i 10 classes d'entre les 20 totals. Com es pot comprovar, si ens fixem amb la precisió obtinguda a la taula 4, s'observa com les dades finals van resultar semblants en relació a l'obtingut amb 20 classes, mantenint aquest 3% de marge. Els resultats finals detallats i les matrius de confusions d'aquests experiments, es poden trobar als apèndix A2 i A3.

TAULA 3

COMPARATIVA DE RESULTATS REDUCTOR SOROLL VS. IMATGES ORIGINALS, EXTRACTOR SURF AMB 20 CLASSES

Mètode	Accuracy (%)
SURF + Imatges originals	66.6
SURF + Imatges soroll reduït	69.6

TAULA 4

COMPARATIVA DE RESULTATS REDUCTOR SOROLL VS. IMATGES ORIGINALS, EXTRACTOR SURF AMB 5 I 10 CLASSES

5 classes + SURF	Accuracy (%)
Data set original	71
Data set amb reducció soroll	74
10 classes + SURF	Accuracy (%)
Data set original	69.5
Data set amb reducció soroll	72.5



(a)

(b)

Fig. 10. Exemple de comptadors visualment semblants. (a) Model de comptador pertanyent a la classe 94 (b) Model de comptador pertanyent a la classe 89.

## 5.2 Experiments amb altres detectors i descriptors

Prendrem a partir d'ara el data set amb les imatges retallades, que són les que millors resultats ens han propor-

cionat. Tot seguit experimentarem entre l'extractor SURF, el detector Dense i l'extractor SIFT que estan entre els 3 models més utilitzats a l'estat de l'art. Donat que el mètode de creació de vocabulari serà diferent a cada una de les subseccions l'haurèm d'anar recreant per a cada un dels experiments.

### 5.2.1 Extractor SIFT

Aquest experiment proposa utilitzar el data set que ens ha donat millors resultats a la secció 5.1 però canviant l'enfoc amb el que s'extreuen les característiques del conjunt d'imatges.

TAULA 5  
COMPARATIVA RESULTATS SURF VS. SIFT

Extractor	Accuracy (%)
SURF	69.6
SIFT	66.2

Els resultats que ens ha proporcionat l'extractor SIFT envers els anteriors obtinguts amb SURF, es poden observar a la taula 5.

En termes generals els resultats obtinguts amb l'extractor SURF segueixen sent un 3% més precisos. Comparant les dues matrius de confusió, però, s'observa una mateixa tendència a l'hora de classificar erròniament comptadors visualment semblants (mateixa marca diferent model). Amb aquest experiment, les classes confoses varien respecte l'experiment amb SURF. Recordem que aquest confonia 8 de les 16 imatges corresponents a la classe 94 amb la classe 89, per exemple. L'enfoc amb SIFT classifica aquesta classe molt millor amb 14 classificacions correctes per només 2 errònies, de les quals només 1 a la classe 89. Aquesta diferència és deguda a la manera de detectar i descriure els punts d'interès com s'explica durant les seccions 4.1.1 i 4.1.2. En canvi falla a l'hora de classificar altres classes que no fallaven amb l'enfoc SURF. Com succeeix amb la classe 75, que classifica només 4 de les 12 imatges de text correctament, 5 a la classe 73 i 3 a la 62. La fig. 11 torna a mostrar com, efectivament, es manté una tendència a confondre comptadors visualment semblants.



(a)

(b)

(c)

Fig. 11. Exemple de comptadors visualment semblants. (a) Model de comptador pertanyent a la classe 75 (b) Model de comptador pertanyent a la classe 73 (c) Model de comptador pertanyent a la classe 62.

### 5.2.2 Dense + SURF / SIFT

El següent experiment va consistir en realitzar una combinació entre el detector Dense i els descriptors SURF i SIFT, independentment. Aquest primer enfoc amb el detector Dense, utilitza els paràmetres d'entrada per defecte. Els paràmetres per defecte del detector Dense es mostren i s'expliquen a l'apèndix A4.

A la taula 6 es mostren tots els resultats obtinguts experimentant amb aquest detector.

El primer enfoc Dense/SURF ens va deixar un sorprenent 5% a la precisió final i un 24% a la precisió a l'hora de classificar el training. D'una manera semblant, una configuració Dense/SIFT va resultar en un 10% d'encert general i un 43% classificant el training. Com ja s'ha comentat un vocabulari nou va ser creat per a cada experiment.

La següent subsecció explora el comportament dels paràmetres d'entrada d'aquest detector.

### 5.2.3 Paràmetres detector Dense

La primera prova va consistir en afegir una i dues escales al detector i observar quins dels dos experiments proporcionen millors resultats.

En relació a l'execució anterior amb paràmetres per defecte, s'aprecia una millora considerable en quant a eficiència en la precisió del training i la global. Destacar que amb el cas de dues escales la millora de la precisió a l'hora de classificar el training és d'un 50% gairebé i la millora global és d'un 20%.

L'apèndix A6 mostra quins han estat els valors dels paràmetres del detector Dense que s'han fet servir en aquest primer experiment així com els resultats finals obtinguts. L'apèndix A4 explica el significat d'aquestes variables.

Prenent com a millor configuració en aquest cas la de les dues escales és va fer una altra prova modificant el paràmetre `initFeatureScale`, el qual defineix el valor de la primera escala.

La primera execució es va fer prenent un valor menor (de 16.f a 8.f) i la segona un de major (de 16.f a 32.f), obtenint millors resultats només amb la segona prova. Per un valor de inici d'escala de 32.f és va obtenir un valor de precisió a la classificació del training del 83% i un 32% global. Aquesta serà la millor eficiència obtinguda amb aquest tipus de configuració detector/descriptor (Dense/SURF).

Amb aquest millor valor es va realitzar un darrer experiment afegint 2 escales més. L'execució amb 4 escales i un valor inicial de 32.f va produir uns resultats finals de 14% de precisió en la classificació final, valors clarament inferiors als anteriors.

TAULA 6

COMPARATIVA RESULTATS AMB DETECTOR DENSE SEGONS DIFERENTS PARÀMETRES I DESCRIPTORS

1 escala	Accuracy (%)
SIFT	10
SURF	5
Params (1) + SURF	9
2 escales	Accuracy (%)
Params (2) + SURF	25.7
Params (3) + SURF	13
Params (4) + SURF	32
4 escales	Accuracy (%)
Params (5) + SURF	14

(1) *Apendix A6 una escala.*

(2) *Apèndix A6 dues escales.*

(3) *initFeatureScale = 8.f, apèndix A6 dues escales.*

(4) *initFeatureScale = 32.f, apèndix A6 dues escales.*

(5) *initFeatureScale = 32.f, featureScaleLevels = 4.*

### 5.3 Tamany de vocabulari

Com a últim experiment i prenent la configuració obtinguda a la secció 5.1 com la més efectiva, es tracta d'observar que succeeix quan fem servir un vocabulari definitament més gran. Fins ara utilitzàvem una mida de 999 descriptors, que venen definits pel número de clústers definits pel k-means (com s'explica durant la secció 4.2). S'han realitzat experiments mitjançant l'extractor SURF amb una mida de vocabulari de 2000 descriptors i un altre experiment amb 4000 descriptors, de manera independent.

A l'apèndix A3 es mostren els resultats detallats i la matriu de confusió pel primer experiment amb 2000 clústers. A la taula 7 es pot observar com s'obtenen pitjors resultats: un 65% contra el casi 70% obtingut amb la millor configuració.

La prova amb vocabulari de 4000 descriptors ha donat resultats casi idèntics a l'anterior amb un 64,8% de precisió global (taula 7). En aquest sentit, es pot percebre una tendència a mantenir l'eficiència per més clústers que hi definim.

TAULA 7

COMPARATIVA RESULTATS DIFERENT MIDA VOCABULARI AMB EXTRACTOR SURF

Vocabulari (num. paraules)	Accuracy (%)
999	69.6
2000	65
4000	64.8

## 6 CONCLUSIÓ

Es pot afirmar que l'objectiu que es va plantejar a l'inici d'aquest projecte ha estat assolit. S'ha experimentat amb algunes de les principals configuracions més utilitzades a l'estat de l'art mitjançant un data set consistent en 20 classes i 1183 imatges i obtenint gairebé un 70% de precisió amb la millor configuració trobada. Els resultats obtinguts amb l'extractor SURF són clarament superiors als obtinguts aplicant el detector Dense i ostensiblement millors als obtinguts amb l'extractor SIFT. S'han examinat els paràmetres configurables del detector Dense. Addicionalment, s'ha aportat una eina de reducció de soroll a les imatges que millora un 3% la precisió del classificador respecte els resultats obtinguts amb el conjunt d'imatges original. S'ha comprovat com fer servir un tamany de vocabulari major no implica necessàriament millors resultats.

A partir d'aquí s'obren moltes possibilitats per a treballs futurs. Trobar una manera d'encarar el problema en la detecció de punts entre comptadors amb mateixa marca i diferent model, pot ser un dels problemes més desafiants a resoldre.

Un altre punt seria el de millorar la robustesa del reductor de soroll proposat en aquest treball. Una possible solució podria estar encaminada a millorar el tractament de selecció de les bounding boxes. També es podria experimentar més amb el valor de lllindar aplicat a l'hora de detectar contorns; s'hauria d'intentar trobar un equilibri en aquest sentit que dongués millors resultats.

Per millorar els resultats fent servir el detector Dense, es podria mirar d'emparellar amb els descriptors SIFT. La secció 5.2.2 mostra uns millors resultats que la parella Dense/SURF amb una escala, però per limitacions produïdes per la manca de temps es va decidir utilitzar el descriptor SURF als experiments posteriors.

Una proposta de solució per trobar un tamany òptim de vocabulari pot ser preprocessar les dades usant un algorisme d'anàlisi de components com PCA [13] i aplicar k-means sobre aquest espai resultant. L'objectiu final seria el de provar de facilitar la feina al k-means per que pugui diferenciar millor els clústers.

## AGRAÏMENTS

Voldria agrair els ànims i empena rebuts per part de la meva família i amics durant la realització del treball.

Al meu tutor, Ernest Valveny, tant pel temps com la paciència dedicades. Gràcies al seu ajut i idees suggerides he pogut incrementar la qualitat final d'aquest projecte.

## BIBLIOGRAFIA

- [1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. *Visual categorization with bags of keypoints*. In Proc. ECCV Workshop on Statistical Learning in Computer Vision, 2004.
- [2] L. Zhu, A. Rao and A. Zhang. *Theory of Keyblock-based image retrieval*. ACM Transactions on Information Systems, **20**, (2), 224-257, 2002.
- [3] C. Harris and M. Stephens. *A combined corner and edge detector*. In Alvey vision conference, volume 15, page 50. Manchester,

UK, 1988.

- [4] T. Tuytelaars. *Dense Interest Points*. K.U.Leuven, ESAT - PSI.
- [5] D. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- [6] H. Bay, T. Tuytelaars, and L. V. Gool. *SURF: Speeded Up Robust Features*. In ECCV (1), 2006, pp. 404-417.
- [7] A. Alahi, R. Ortiz and P. Vandergheynst. *FREAK: Fast Retina Keypoint*. Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
- [8] "OpenCV Documentation". <http://docs.opencv.org> (Últim accés el 19/06/2015).
- [9] O. Duda, P.E. Hart, D.G. Stork. *Pattern classification*. John Wiley & Sons, 2000.
- [10] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- [11] "Support Vector Machines" (SVM). <http://www.statsoft.com/Textbook/Support-Vector-Machines> (Accedit el 22/06/2015).
- [12] D. Douglas and T. Peucker. *Algorithms for the reduction of the number of points required to represent a digitized line or its caricature*. The Canadian Cartographer 10(2), 112-122 (1973).
- [13] I.T. Jolliffe. *Principal Component Analysis*. Springer, 2002.



# APÈNDIX

## A1. MATRIUS CONFUSIÓ EXTRACTOR SURF

TAULA 8  
MC AMB 20 CLASSES SENSE APLICAR REDUCTOR DE SOROLL I EXTRACTOR SURF

Target	53	55	58	62	68	69	70	71	72	73	76	81	82	83	86	88	89	90	91	94
53	7	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
55	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	5
62	0	2	0	2	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0
68	0	0	0	0	8	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0
69	0	0	0	1	0	3	0	0	0	4	0	0	0	0	0	0	1	0	0	0
70	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	10	0	0	0	2	0	0	0	0	0	0	0	0	0
72	0	1	0	0	0	0	0	10	0	0	2	0	0	0	0	0	0	0	0	0
73	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0
76	0	1	0	1	0	0	0	0	2	7	0	0	0	0	0	0	1	0	0	1
81	0	0	0	0	0	0	1	0	0	0	11	0	0	0	0	0	0	0	0	0
82	0	0	0	0	0	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0
83	0	1	0	0	0	0	0	0	0	0	0	0	7	0	2	2	0	0	1	1
86	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	1	1
88	0	0	0	0	0	0	0	0	0	0	0	0	2	0	7	3	0	0	1	1
89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	1	1
90	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	2	7	1	0	0
91	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	9	0
94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	7	7

-> Training descriptors: 946

-> Training Classification Accuracy = 0.966173

-> Overall Classification Accuracy = 0.666667

TAULA 9  
MC AMB 20 CLASSES APLICANT REDUCTOR DE SOROLL I EXTRACTOR SURF

Target	53	55	58	62	68	69	70	71	72	73	76	81	82	83	86	88	89	90	91	94
53	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
58	0	5	4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2
62	0	2	0	3	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
68	0	0	0	0	6	0	0	0	4	0	2	0	0	0	0	0	0	0	0	0
69	0	0	0	0	3	0	0	0	5	0	0	0	0	0	0	1	0	0	0	0
70	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	11	0	0	1	0	0	0	0	0	0	0	0	0	0
72	1	0	0	0	0	0	0	8	0	4	0	0	0	0	0	0	0	0	0	0
73	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0
76	0	0	0	0	0	0	0	0	1	9	0	0	1	0	0	2	0	0	0	0
81	0	0	0	0	0	0	1	0	0	0	11	0	0	0	0	0	0	0	0	0
82	0	0	0	0	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0	0
83	0	0	0	0	0	0	0	0	0	0	0	11	0	1	0	1	0	0	0	1
86	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	1	1
88	0	0	0	0	0	0	0	0	0	0	0	4	0	6	1	0	0	2	0	2
89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13	0	0	1	1
90	2	1	0	0	0	0	1	0	0	0	0	0	0	1	0	1	7	1	0	0
91	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9	0	0
94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	8	8

-> Training descriptors: 946

-> Training Classification Accuracy = 0.975687

-> Overall Classification Accuracy = 0.696203

## A2. MC I RESULTATS SENSE APLICAR REDUCTOR I EXTRACTOR SURF

TAULA 10  
MC I RESULTATS AMB 5 CLASSES

Target	88	89	90	91	94
88	9	3	0	0	1
89	0	14	0	0	1
90	0	3	10	1	0
91	0	1	1	9	0
94	0	9	0	0	7

-> Training descriptors: 273

-> Training Classification Accuracy = 0.985348

-> Overall Classification Accuracy = 0.710145

TAULA 11  
MC I RESULTATS AMB 10 CLASSES

Target	75	81	82	83	86	88	89	90	91	94
75	9	0	0	1	0	1	1	0	0	1
81	0	12	0	0	0	0	0	0	0	0
82	0	6	4	0	0	0	0	0	0	0
83	0	0	0	8	0	2	2	0	0	1
86	0	0	0	0	12	0	1	0	0	1
88	0	0	0	2	0	7	3	0	0	1
89	0	0	0	0	0	0	14	0	0	1
90	0	0	0	0	1	0	3	9	1	0
91	0	0	0	0	0	0	1	1	9	0
94	0	0	0	0	0	0	9	0	0	7

-> Training descriptors: 521

-> Training Classification Accuracy = 0.984645

-> Overall Classification Accuracy = 0.694656

## A3. MC I RESULTATS APLICANT REDUCTOR I EXTRACTOR SURF

TAULA 12  
MC I RESULTATS AMB 5 CLASSES

Target	88	89	90	91	94
88	9	2	0	0	2
89	0	14	0	0	1
90	0	2	11	1	0
91	0	2	0	9	0
94	0	8	0	0	8

-> Training descriptors: 273

-> Training Classification Accuracy = 0.985348

-> Overall Classification Accuracy = 0.73913

TAULA 13  
MC I RESULTATS AMB 10 CLASSES

Target	75	81	82	83	86	88	89	90	91	94
75	10	0	0	1	0	0	2	0	0	0
81	0	12	0	0	0	0	0	0	0	0
82	0	5	4	0	1	0	0	0	0	0
83	0	0	0	11	0	1	0	0	0	1
86	0	0	0	0	12	0	1	0	0	1
88	0	0	0	4	0	6	1	0	0	2
89	0	0	0	0	0	1	13	0	0	1
90	1	0	0	0	1	0	1	10	1	0
91	0	0	0	0	0	0	2	0	9	0
94	0	0	0	0	0	0	8	0	0	8

-> Training descriptors: 521

-> Training Classification Accuracy = 0.986564

-> Overall Classification Accuracy = 0.725191

## A4. DEFINICIÓ DETECTOR DENSE

Definició de la funció DenseFeatureDetector amb els seus paràmetres per defecte:

**DenseFeatureDetector**( float **initFeatureScale**=1.f, int **featureScaleLevels**=1, float **featureScaleMul**=0.1f, int **initXyStep**=6, int **initImgBound**=0, bool **varyXyStepWithScale**=true, bool **varyImgBoundWithScale**=false )

- **initFeatureScale** - Defineix el valor de la primera escala.
- **featureScaleLevels** - Determina el número de nivells d'escala a utilitzar.
- **featureScaleMul** - Aquest parametre indica el valor pel que es multipliquen els parametres que defineixen les característiques de cada nivel (**initFeatureScale**, **initXyStep** i **initImgBound**)
- **initXyStep** - Defineix quin valor te el tamany d'un node a una escala.
- **varyXyStepWithScale** - Flag per decidir si es multiplica el valor de **initXyStep** a cada escala pel valor **featureScaleMul**.
- **varyImgBoundWithScale** - Flag per decidir si es multiplica el valor de **initImgBound** a cada escala pel valor **featureScaleMul**.

## A5. MC I RESULTATS AMB UN VOCABULARI DE 2000 CLÚSTERS I EXTRACTOR SURF

TAULA 14  
MC I RESULTATS AMB 20 CLASSES, EXTRACTOR SIFT I 2000 CLÚSTERS

Target	53	55	58	62	68	69	70	71	72	73	75	81	82	83	86	88	89	90	91	94	
53	6	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0
55	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
58	0	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
62	0	1	1	3	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
68	0	0	0	0	7	0	1	0	0	4	0	0	0	0	0	0	0	0	0	0	0
69	0	0	0	0	0	2	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	0	9	0	0	0	3	0	0	0	0	0	0	0	0	0
72	0	0	0	0	0	0	0	0	10	0	0	2	0	0	0	0	1	0	0	0	0
73	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0
75	0	0	0	1	0	0	0	0	0	2	8	0	0	0	0	0	2	0	0	0	0
81	0	0	0	0	0	0	0	1	0	0	0	11	0	0	0	0	0	0	0	0	0
82	0	0	0	0	0	0	0	0	0	7	0	0	3	0	0	0	0	0	0	0	0
83	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	1	0	0	2	
86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	1	
88	0	1	0	0	0	0	0	0	0	0	0	0	0	5	0	4	1	0	0	2	
89	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	13	0	0	0	
90	3	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	5	2	1	
91	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1
94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	7	

-> Training descriptors: 946

-> Training Classification Accuracy = 0.973573

-> Overall Classification Accuracy = 0.649789

## A6. PARÀMETRES D'ENTRADA DETECTOR DENSE AMB 1 I 2 ESCALES I DESCRIPTOR SURF.

Amb **una escala** es va fer servir la següent configuració de paràmetres:

```
DenseFeatureDetector( float initFeatureScale=8.f, int featureScaleLevels=1, float featureScaleMul=0.5f,
int initXyStep=16, int initImgBound=0,
bool varyXyStepWithScale=true, bool varyImgBoundWithScale=false )
```

Obtenint els resultats finals següents:

Training descriptors: 946

Training Classification Accuracy = 0.541226

Overall Classification Accuracy = 0.0886076

Amb **dues escales** es va fer servir la següent configuració de paràmetres:

```
DenseFeatureDetector( float initFeatureScale=16.f, int featureScaleLevels=2, float featureScaleMul=0.5f,
int initXyStep=32, int initImgBound=0,
bool varyXyStepWithScale=true, bool varyImgBoundWithScale=false )
```

Obtenint els resultats finals següents:

Training descriptors: 946

Training Classification Accuracy = 0.711416

Overall Classification Accuracy = 0.257384