



**EPS**

Escola Politècnica  
Superior

## Projecte/Treball Fi de Carrera

**Estudi:** Eng. Tècn. Informàtica de Gestió. Pla 2001

**Títol:** Reconeixement de CD-ROMS musicals mitjançant visió per computador

**Document:** Resum

**Alumne:** Eloy Roura Perez

**Director/Tutor:** Jordi Freixenet Bosch; Xavier Lladó Bardera

**Departament:** Arquitectura i Tecnologia de Computadors

**Àrea:** Arquitectura i Tecnologia de Computadors

**Convocatòria** (mes/any): juny 2009

## 1. Introducció

En aquest treball final de carrera hem decidit centrar-nos en el camp de visió per computador. Aquest camp és molt ampli i en el nostre cas tractarem el reconeixement d'imatges a partir d'una aplicació divertida. Concretament analitzarem caràtules de cd-roms, les quals haurem de reconèixer per després reproduir automàticament la música del seu àlbum associat.

Amb aquest projecte es vol assolir l'objectiu de dissenyar i desenvolupar un classificador d'imatges. Per tant del que es tracta és d'estudiar tècniques que permetin extreure característiques així com descriure les imatges. També s'analitzaran tècniques que permetin utilitzar aquesta informació extreta d'un conjunt d'imatges per aprendre a reconèixer i a classificar-ne de noves.

Per tant podríem definir l'objectiu principal com:

**Dissenyar, implementar i testejar un sistema per classificar imatges:** Volem dissenyar un sistema que primer aprèn com són les imatges d'una classe a partir d'un conjunt d'imatges d'entrenament i després és capaç de classificar noves imatges assignant-els-hi la etiqueta corresponent a una de les classes "apreses".

Per assolir aquest objectiu ens n'hem proposat d'altres com:

- Generar un sistema de visió per computador on extraurem característiques, entrenarem les imatges amb aquestes característiques i finalment testejarem amb noves imatges.
- Respecte a la part tecnològica, crearem una aplicació on-line amb la qual puguem generar una base de dades d'imatges i música. L'aplicació també ens haurà de permetre aprendre a descriure aquestes imatges així com capturar-ne de noves per tal de testejar el sistema de reconeixement.

## 2. Reconeixement de cd-roms

Per desenvolupar el sistema que hem descrit haurem de dividir-lo en dues etapes, la primera servirà perquè el classificador conegui totes les caràtules que tenim (fase d'entrenament) i la segona ens servirà per testejar les imatges (fase de classificació). En la figura 1 veiem un esquema de les fases que descrivim.

La fase 1 de l'esquema consisteix en descriure les imatges originals. Per obtenir una caracterització d'aquestes imatges utilitzarem SIFT (Scale Invariant Features Transform) que permet extreure un vector de característiques per a cada punt d'interès de la imatge. Un cop es tenen els descriptors de SIFT haurem de guardar a la base de dades tota la informació de l'entrenament per tal que després puguem fer la classificació. Imatges originals, música i descriptors de les imatges s'emmagatzema a la base de dades.

En la fase 2, el classificador ja coneix totes les classes, només falta crear l'estructura que es farà servir per classificar les noves caràtules capturades. Un cop es reconeix a quin cd pertany la nova caràtula, automàticament es busca a la base de dades la música associada i es reproduïx. La captura de noves imatges de test la farem a partir d'una webcam i la reproducció es farà en un software extern a l'aplicació.

Per generar el classificador s'han analitzat diferents estratègies, des de les més bàsiques i intuïtives fins a d'altres de més complexes. La inicial va ser l'estratègia de distància mínima usant la tècnica del veí més proper, on comparàvem descriptor a descriptor donant un vot a cada descriptor amb distància mínima. Degut a que aquesta tècnica, a priori pot resultar molt costosa hem pensat en altres estratègies basades en arbres de cerca. Dins les estratègies basades en arbres s'han provat diferents implementacions possibles, totes relacionades amb la forma de voteig i les mides de l'arbre. Aquestes diferents estratègies s'han avaluat en termes de classificació d'imatges i cost computacional en els experiments d'aquest PFC.

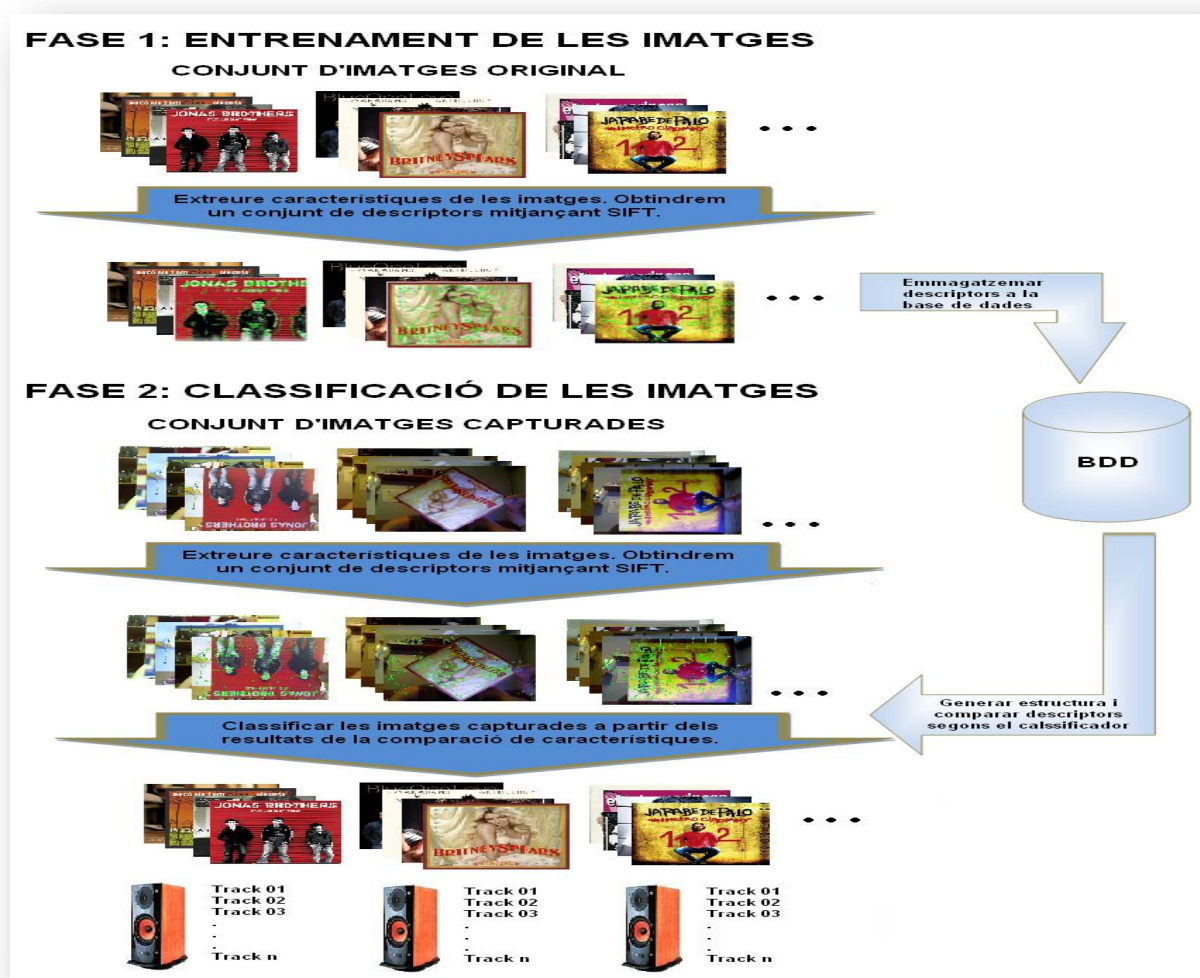


Figura 1: Esquema del sistema dissenyat.

### 3. Resultats

En el desenvolupament del sistema hem proposat diferents estratègies de classificació que després hem analitzat avaluant els resultats. Hem començat amb els mètodes més senzills, com són distància mínima entre els descriptors de les imatges i comparant tots amb tots i hem acabat generant estructures més complexes basades en arbres de cerca per fer la classificació més ràpidament.

Per dur a terme aquests experiments hem treballat amb un conjunt de classes (caràtules de cd-roms) reduït de 10 cd-roms al iniciar el prototip i un altre conjunt de 100 cd-roms quan l'hem refinat. Aquestes imatges originals i utilitzades per l'etapa

d'entrenament les hem descarregat totes des de la pàgina Amazon. A l'hora de fer el test del sistema hem capturat fins a 1600 imatges diferents. Al inici hem avaluat imatges amb poca variabilitat, estudiant rotacions amb pocs canvis d'escala (distància del cd a la càmera). Aquest primer conjunt estava compost per 600 imatges de test. Llavors, després de fer els primers tests vam capturar un nou conjunt de 400 imatges amb un canvi d'escala més elevat. I finalment vam arribar a un grup de 300 imatges aplicant tot tipus de variacions (escala, rotacions, canvis de perspectiva, canvis d'iluminació...).

En la figura 2 veiem el conjunt original de 100 cd-roms recollides de l'Amazon i utilitzades per l'entrenament.



Figura 2: Conjunt de 100 imatges descarregades de Amazon.

Amb la primera estratègia descrita, hem iniciat els testos amb un subconjunt de la figura 2 de 10 cd-roms per entrenar i 60 imatges per testejar. Des d'aquest punt cada vegada que hem assolit un objectiu hem incrementat la dificultat del problema amb noves imatges. Així hem arribat a tenir un més d'un 90% en aquestes primeres avaluacions. Més endavant quan hem avaluat les estructures d'arbre de cerca hem vist com milloraven els resultats. Des de l'inici fins al final hem passat d'un 90% fins a un

99% d'encert en el conjunt d'imatges amb poca variabilitat. Cal destacar un fet important, en la primera estratègia trigàvem més de 6h a testear tot el conjunt de 600 imatges, temps que s'ha reduït considerablement utilitzant un classificador basat en arbre, quedant-se en només 13 minuts per totes les 600 imatges. L'únic que hem empitjorat i no excessivament ha sigut el temps de creació de l'estructura per al classificador, que ha passat de menys de 5 segons a 38 segons.

Per tant, com és evident, la millor opció per al nostre problema de reconeixement de cd-roms és la del classificador basat en un arbre de cerca.

## 4. Conclusions

Finalment hem sigut capaços de construir un sistema de classificació d'imatges més que satisfactori. Arribant a tenir una base de dades de 100 caràtules i tenir un percentatge d'encert aproximadament del 99% en un conjunt de test de 100 imatges que arriben a ocupar fins a  $\frac{1}{4}$  de la imatge, i que poden estar rotades 90 graus, 180 graus... En canvi vam fer proves amb un conjunt d'imatges amb més variabilitat com canvis de perspectiva, augment de canvis d'escala... i el resultat va ser del 79%.

Com a contribucions d'aquest PFC:

- Hem dut a terme un anàlisi, disseny i implementació d'un sistema de reconeixement d'objectes en imatges. Hem usat diferents alternatives com són: el mètode del veí més proper i altres estratègies basades en construcció d'arbres de cerca.
- Hem sigut capaços de crear una aplicació per poder integrar tota la implementació del sistema. Aquesta aplicació ens permet la captura d'imatges (caràtules de cd-roms), i a partir del seu reconeixement ens fa sonar l'àlbum pertinent.
- Hem creat un conjunt d'imatges força ampli que està al abast de tothom. La base de dades consta de 1600 imatges capturades i 100 més d'originals.