

**EPS**Escola Politècnica
Superior

Projecte/Treball Fi de Carrera

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

Títol: Reconeixement d'escenes de la ciutat de Girona a partir del contingut d'imatges

Document: Memòria

Alumne: Carles Fontanella Bonjoch

Director/Tutor: Jordi Freixenet / Xavier Lladó

Departament: Arquitectura i Tecnologia de Computadors

Àrea: Arquitectura i Tecnologia de Computadors

Convocatòria (mes/any): 06/2009

1. Idea principal

El PFC s'ha realitzat en el departament d'Arquitectura i Tecnologia de Computadors (ATC) de la Universitat de Girona dins del grup de Visió per Computador i Robòtica, www.vicorob.udg.edu, seguint la línia de recerca que estudia el reconeixement d'objectes dins d'una imatge. La idea principal de projecte és realitzar una aplicació que permeti reconèixer els objectes que apareixen en una imatge per llavors donar informació referent al contingut de l'imatge a l'usuari que ha fet la consulta.

2. Objectius

Els objectius del PFC consisteixen en desenvolupar una aplicació informàtica basada en un sistema de visió per computador, la qual permeti donar una resposta en forma d'informació a partir d'una *query* d'una imatge que conté una escena o objecte en concret. Es pot veure un esquema global d'aquest procés a la Figura 1



Figura 1 Exemple del procés que segueix l'aplicació. Donada una imatge d'entrada, l'aplicació l'analitza, la descriu i la classifica i finalment retorna informació sobre l'objecte que hi apareix.

A l'hora de descriure una imatge segons les seves característiques poden aparèixer molts problemes relacionats amb l'orientació dels objectes que hi apareixen, l'escala, l'il·luminació de la fotografia, etc. El nostre sistema tindrà en compte aquests factors i tractarà de resoldre aquests problemes amb els mètodes més adequats.

Així, doncs, el principal objectiu del PFC és analitzar, dissenyar i construir un sistema de visió per computador capaç de reconèixer objectes d'interès en imatges.

3. Descripció de les imatges

Tot el procés segueix una estratègia de reconeixement basada en Bag of Words. Aquest procés es pot desglossar en 3 etapes.

1. Construcció del vocabulari

Primer de tot, s'extreuen les principals característiques (keypoints) de les imatges mitjançant el mètode SIFT(Scale-Invariant Feature Transform). La idea és identificar aquests descriptors que caracteritzen a una sèrie d'imatges prototipus per llavors extreure'n les paraules visuals representatives, construir un vocabulari i llavors utilitzar aquest vocabulari per descriure les imatges tant del conjunt d'entrenament com el de test. Es pot veure un exemple de la creació del vocabulari visual a la Figura 2.

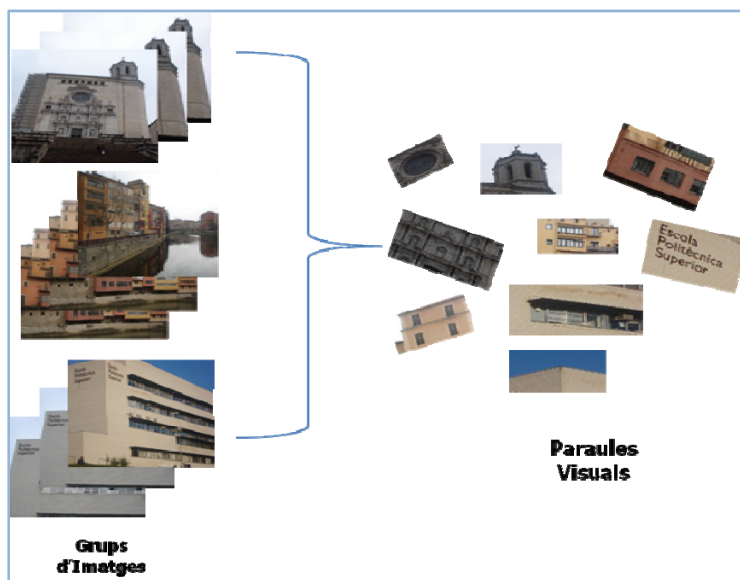


Figura 2 Exemple de creació d'un vocabulari. D'un conjunt d'imatges de 3 classes diferents s'extreuen un determinat número de paraules visuals. Aquest vocabulari ens servirà per descriure les imatges mitjançant histogrames.

Un cop es té aquest vocabulari, el que es fa és descriure amb histogrames la resta d'imatges i anar comptabilitzant el nombre de paraules visuals de cada tipus que es localitzen a cada una, per llavors poder saber a quina classe (tipus d'imatge) correspon o amb quina classe està més relacionada. Aquests histogrames tindran a l'eix de les X totes les paraules visuals del conjunt, i a l'eix de les Y el nombre de vegades que apareixen en aquella imatge, tal com mostra l'exemple de la Figura 3.



Figura 3 En aquest exemple, s'utilitza un vocabulari de 4 paraules visuals. Es descriu l'imatge de la dreta segons aquest vocabulari i es pot comprovar com apareix algunes vegades les paraules visuals de "mar", "gespa" i "cel" però no la del "cotxe".

2. Aprenentatge

Fins al moment hem plantejat com descriure les imatges a partir d'un vocabulari visual. A continuació ens caldrà un procés d'entrenament, és a dir, un procés d'aprenentatge en el qual el sistema aprèn de forma automàtica a descriure i reconèixer els objectes d'interès d'una determinada imatge. Així doncs, ens farà falta un conjunt d'imatges d'entrenament (on coneixem cada imatge quina classe pertany), que descriurem fent servir les paraules visuals que s'han definit en l'anterior apartat, i que utilitzarem per construir un sistema classificador. Com que tenim les imatges correctament etiquetades, s'agafarà un número d'imatges de cada grup per fer l'entrenament del sistema. En aquest PFC s'utilitzaran diferents classificadors (LDC, Fisher, SVM, KNN) per fer l'entrenament, de manera que podrem donar diferents resultats i comparar i avaluar els diferents mètodes.

3. Testing

En aquesta part, donada una imatge de query de la qual no en coneixem res, s'analitza i es construeix el seu 'bag of words' segons el vocabulari visual. Per tant, obtenim un histograma que es podrà comparar amb la nostra base de consulta. Mitjançant, aquest procés obtindrem un resultat de classificació de la nova imatge, tal com es mostra en la Figura 4.

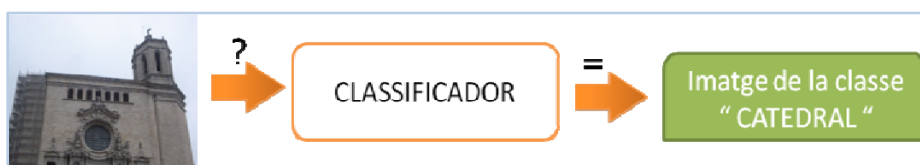


Figura 4 Exemple de consulta d'una imatge al nostre classificador creat. El sistema retorna informació sobre l'objecte d'interès que apareix a l'imatge.

4. Aplicacions desenvolupades

L'aplicació realitzada es pot dividir en dos mòduls:

La primera correspon a l'aplicació principal o aplicació base, amb la qual l'usuari pot analitzar les imatges, descriure-les, obtenir histogrames, entrenar i testear amb elles i construir un sistema el més robust possible, tal com marca els objectius del projecte. Aquest primer mòdul només està orientat al programador o a la persona que vol tractar amb les imatges.

El segon, correspon a la capa superior, la capa que és visible a l'usuari final i que podrà fer ús de l'aplicació dissenyada amb tots els seus avantatges sense tenir, necessàriament, coneixements de visió per computador. Aquest mòdul estarà representat per una pàgina web que simularà un possible funcionament futur de l'aplicació adaptat a un entorn web on qualsevol usuari podrà enviar una imatge i obtenir-ne informació.

5. Resultats

S'han realitzat diferents proves per tal d'analitzar el funcionament de l'aplicació. Les imatges escollides han estat elements de la ciutat de Girona. Es disposa d'una base de dades d'imatges dels principals elements arquitectònics i turístics de la ciutat, a la següent Figura 5, es pot veure alguns exemples.



Figura 5 Petit recull d'imatges de que es disposa.

S'ha començat realitzant provés més senzilles (menys número de classes) i s'ha anat incrementant la dificultat. També s'ha variat el número de paraules visuals utilitzades per crear el vocabulari.

Els resultats de 3 classes han estat bons. El sistema té un alt percentatge d'encerts en general, utilitzant tots els classificadors amb un vocabulari visual de 50 paraules (del 90% fins al 95% d'encerts). Amb 5 classes, el sistema continua mostrant bons resultats. Es comencen a veure els classificadors que aniran millor per a la nostre aplicació : el SVM i el KNNC (del 87% fins al 91%). Amb 7 classes el sistema disminueix lleugerament el percentatge d'encerts (del 82% al 88%). Es pot veure que els resultats són molt millors amb 50 paraules visuals que amb valors superiors. Els resultats amb 8, 9 i 10 classes continua passant el mateix que en el cas de 7. És en aquests experiments on es constata que els millors classificadors són el SVM i el KNNC.

Un cop s'han realitzat els experiments, s'ha pogut veure que el grup d'imatge que s'utilitza per realitzar el vocabulari i la part de *training* influeix molt en el resultat. Si el grup escollit representa bé les principals característiques de cada classe els resultats seran bons. El nombre de paraules visuals que hem vist que millor funciona, en general, en tots els experiments és el de 50. Els classificadors més adequats per tractar les nostres dades han estat el SVM i el KNNC.

6. Conclusions

Després de realitzar el projecte, es pot dir que s'han assolit els objectius que inicialment s'havien establert. S'ha desenvolupat una aplicació capaç de crear un sistema més o menys robust que té en compte factors com la rotació d'objectes dins d'una imatge, escala i il·luminació. A través d'aquesta aplicació és possible crear diferents variants d'aquesta base de consulta fent modificacions, consultant els seus components, així com poder analitzar cada fase del procés per separat. Les principals aportacions d'aquest projecte són:

- Sistema de visió per computador que permet fer la descripció i reconeixement de les imatges
- Demostrador web : Una aplicació web que utilitza el sistema de reconeixement i simula una possible aplicació final d'aquest PFC.
- Col·lecció de 350 fotografies de diferents elements arquitectònics i típics de la ciutat de Girona, preparades per ser analitzades i classificades amb aquesta mateixa aplicació o per ser aprofitades per algun altre objectiu.