Face recognition is a task which is executed on a daily basis, efficiently and without special efforts by any human beeing. However, this doesn’t hold in the computer’s world where a face recognition problem like any other computer vision problem is a non trivial one. Detecting a face from an input image, finding it’s scale and contour, discovering it’s shape parameters and rendering the original face into a standard shape format, finding a way to represent the face data in a low dimmensional space which allows two faces to be compared one to each other, solving the classification problem, all of these are non trivial problems, which require a robust and efificent solution in order to implement any modern face recognition system.

Speaking about terminlogy, the „Face Recognition”, „Face Verification” and „Face Identification” terms are tightly related to each other and they can be easly confused. Face recognition represents the general topic, which includes both face verification and face identification (also reffered to as „face authentication”). On the one hand, face identification is concerned with finding the identity of a face from a set of known identities (one to many matching). The key aspect is that a face identification system knows preciesly the set of identities with which it is dealing, meaning that all the querry faces will be compared against a known set of identities. Obviously in this case, the natural approach is to train the system to learn particular features of the registered identities and search for them in the querry photos. On the other hand, face verification deals with validating a claimed identity of a face, by comaring it with a model face(one to one matching). The system doesn’t have any information about the identities of the faces that are going to be compared so an identity specific training cannot be done here. The key aspect is that the system needs to be general enough, in order to be able to deal with different identities, without loosing its accuracy. It is obvious that a tradeoff needs to be done between flexibility (do we to re-train the system when a new identitfy is added to the system ? ) and accuracy ( how accurate can we be when validating the identity of a person, using only one sample of its face?). Even though we saw that a Face Verification System is different comparing to a Face Identitifcation System, they have a lot things in common. There are a lot of methods (image pre-processing, data representation, classification methods) which are employed by both systems. So even tough the face recognition research field is a wide one, any improvement brought in any level can have wide ramifications.

In the last few years, the research in the face recognition field has focused on five main complementary dirrrections: face detection, landmarks keypoints detection, face registration, face description and face representation using statistical methods. Any modern face recognition system includes all these steps but in general it is enough to improve only one of them in order to increase the overall system performance.

Face Detection

Landmarks detection

Face Registration

Face representation using statistical methods

Invatarea prin mecanisme statistice este in general folosita pentru a mapa fete la encodari reprezantative intr-un spatiu euclidian in care se pot aplica metode de inavatare a diferitor metrici sau metode de cuantificare a similaritatilor. Cercetarile din ultimii ani s-au axat pe descoperirea de noi reprezentari ale imaginilor, sau modolitati de fuzionare a mai multor reprezentari consacrate, cu scopul obtinerii de rezultate mai eficiente fata de modalitatile standard din domeniu (ex SIFT).

In [] se propune o tehnica de fuzionare bazata pe CCA, ce uneste mai multi vectori de caracteristici ai unei imagini intr-un singur vector, ce este mai reprezentativ decat oricare din vectorii primiti la input.

(*descriere cca*)

Fisher vectors

Conceptul de Fisher Kernels reprezinta un framework puternic in domeniul pattern classification. Idee de baza consta in caracterizarea unui semnal primit ca input pe baza unui gradient derivat dintr-un model probabilistic generativ si pasarea rezultatului obtinut la un clasificator discriminativ. Astfel conceptul de Fisher Kernels inglobeaza avantajele a doua abordarari diferite si anume prelucrarea datelor pe baza unui model generativa respectiv clasificarea discriminativa. FK s-a folosit cu succes in domeniul clasificarii de imagini si anume, folosind imagini ca semnal de intrare iar ca model generativ un vocabular visual (Gaussian mixture model) ce aproximeaza distributia caracteristicilor extrase dintr-o imagine.

Fiind dat un vocabular vizual, descrierea unei imagini se reduce la determinarea numarului de apartii al fiecarui cuvant din dictionar, in imaginea primita la intrare. Din acest punct, orice tip de clasificator poate fi folosit pentru categorisierea acestor reprezentari de tip histograma. Majoritate lucrarilor de cercetare s-au concentrat pe determinarea in mod corect a dictionarului vizual. Pentru aceasta s-au folosit mai multe abordari. Ex K-means, Gausssian Mixture Models, mean-shift.

(*descriere K-means*)

(*descriere GMM*)

(*descriere mean-shift*)

S-a observat ca, chiar si in cazul bazelor de date mici cu un numar redus de clase, cele mai bune rezultate s-au obtinut in cazul folosirii un vocabular vizual cu un numar cat mai mare de intrari. Insa aceasta aduce un cost semnifciativ suplimentar, deorece in mod evident, costul computational al calcularii unei histograme depinde in mod direct de numarul de intrari din vocabularul folosit. De aceea s-a cautat folosirea unui vocabular cat mai compact, fara a aduce insa deficituri ale perfomantei. Un exemplu de solutie consta in organizarea vocabularului intr-o structura arborescenta (Randomized Clustering Forests). Alta abordare (information bottleneck principle).

Desi ambele abordari au dat rezultate satisfacatoare ex. Reducerea de 5 ori a dimensiunii, fara scaderea performantei sistemului, vocabularul obtinut nu este unul universal. Ceea ce inseamna ca pentru a dauga noi categorii la sistem, trebuie ca intregul proces de antrenare sa fie refacut.

Intrucat ideea de a avea simultan un vocabular universal si compact pare imposibila, directii din cercetare s-au axat in schimb pe construirea a cate unui vocabular pentru fiecare clasa de obiecte folosita la antrenare. 🡪 K means clustering. Cu toate o abrodare de acest gen este flexibila, costul computational total nu este neaprat micsorat.

Ideea din spatele FK consta in a caracteriza o imagine pe baza unui model pe baza unei functii de probabilitate ce modeleaza procesul de generare al semnalului. Reprezentarile sub forma de gradient ce stau la baza FK au marele avantaj fata de BOV ca sunt mult mai mari dpdv al dimensiunii ceea ce permite folosirea de clasificatoare liniare.

O decizie importanta ce se impune a fi luata in design-ul de modelui generativ consta in folosirea sau nu a claselor, mai precis, alegerea intre invatarea vocabularului intr-un mod supervizat sau ne-supervizat. Cu sigurnta alegerea unei modalitati supervizate va aduce rezultate mai bune insa scade flexibilitatea sistemului. Adaugarea unei noi clase va impune antrenarea intregului sistem.

Tehnicile prezente de pattern classification se impart in 2 directii distincte: abordari generative si abordari discriminative. In timp ce prima abordare, se axeaza pe modelarea de functii probabilistice de densitate, parametrizate in functie de clasa obiectelor clasificte, cea de a doua abordare se axeaza direct pe problema clasificarii. Cu toate acestea, abordarile generative au un numar de proprietati care le fac inca atractive de exemplu, posibilitatea de a utiliza date cu lunigimi variabile.

Fisher kernels au fost concepute cu scopul de a combina beneficiile aborariilor de tip discriminativ cu cele ale abordarilor de tip generativ.

Fie p functie probabilistica de densitate si fie

Fisher kernels principles

Ideea de baza consta in impartirea imaginii primite la intrare in patch-uri multiple, ce se suprapun si extragerea de caracteristici (ex SIFT) pentru fiecare patch in parte. Caracateristicile extrase sunt apoi encodate intr-un singur vector de caracteristici. Cea mai populara modalitate de encodare este BOVW (Bag of visual words) [] (*descriere BOVW*)

Cea mai populara abordare

<http://www.xrce.xerox.com/content/download/6995/52427/file/2006-034.pdf>

3D reconstruction is the process of creating a three dimmensional model from a set of faces captured in different poses. It is the inverse process of capturing a photo from the real world. From a geometrical point of view, an image is the projection of an object represented in a three dimmensional space, into a two dimmensional space. After the projection is done, any depth information is lost. So, at a further point, it’s almost imposbile to recover it using just a single image. This is caused by the fact that for any point in the input image it’s depth stays on a projection ray which is a line ortogonal to the image space. Thus, the depth finding problem has infinite valid solutins. However if two or more images ar available it is possible to determine preciesly the 3D position of a point. Using multiple views, the depth of a point can be determined by tracing all the projection rays (one for each view) and interstecting them. The problem might not have a valid solution in which case the optimal intersection point is determined. In speciality terms, this process is also known as triangulation. The principle behind it is based on the fact that multiple set of points are still related one to each other even when a small number of parameters of the model that describes them are changed. And even more, additional information can be extracted by analyzing how this relationship holds in different situations.

The task of estimating a 3D model consists of a series of processing steps. The fundamental step is represented by the camera calibration process also known as camera resectioning process. The goal is to determine the intrinsic and extrinsic parameters of a pinhole camera which could have been used to project the original 3D model into the current two dimmensional shape. The standard representation systems describe the camera paramters using a 3x4 matrix, so called the camera matrix.