

Паралелни алгоритми за откриване и разпознаване на лица

Кирил Алексиев, Павлина Константинова, Людмил Божилов

Parallel Algorithms for Face Detection and Recognition Kiril Alexiev, Pavlina Konstantinova, Ljudmil Bojilov

Увод

В съвременния свят се наблюдава огромно нарастване на цифровата видеоинформация. Понастоящем тя ни залива отвсякъде – телевизия, интернет, милиони видеокамери, работещи 24 часа на ден. Добре известният разузнавателен безпилотен самолет Global Hawk предоставя видеоинформация от поне три сензора (CCD камера Kodak, инфрачервена камера и SAR сензор) от един свой полет за събитията в област от 40000 квадратни морски мили с разрешаваща способност 90 см. Камерите на неговия по-малък и по-евтин колега Predator могат да разчетат текст с големина 7.5 – 15 см.

Единственият начин да се справим с този информационен бум е да възложим все по-голям кръг задачи на компютрите. Автоматичната обработка на тази информация е сериозно предизвикателство поради три основни причини.

Първата от тях е огромното количество информация, генерирана даже от един единствен видеосензор. Нека за пример вземем модерен телевизионен приемник в стандарта на HDTV, чиито кадри са с размери 1920×1080. Т.е един кадър (фрейм) съдържа информация за 2073600 точки в трите цвята (някои приемници работят даже с 4 основни цвята). Ако за кодиране на цвета се използва 1 байт, информацията за един фрейм изисква около 6.3 Mb. Честотата на обновяване на фреймовете е 25 Hz. В заключение се получава, че информацията, която постъпва само от един телевизионен канал за една секунда, е от порядъка на 157.5 Mb.

Втората причина се крие в сложността на алгоритмите за обработка на видео информацията. Обработката започва с прилагане различни видове филтри за отстраняване на типичния за видеосензорите шум. Следва търсене на различни видове характерни особености в изображението, без които е невъзможно откриване на обекти, разкриване на взаимовръзката им в сцената и др. Търсенето на динамиката на промяна на сцената (междуфреймова обработка), следене на обекти, оценяване на тяхното поведение, прогнозиране на бъдещи събития, откриване на нови събития и др. е също така изключително сложно поради факта, че обектите не са точкови, променят си размера и конфигурацията, често пъти се засенчват и т.н.

Третата причина се крие в това, че много от дейностите, които човек извършва с лекота, наблюдавайки информация от видеосензор, даже не са алгоритмизирани.

Към тези три причини може да се добави и изискването за получаване на търсената информация в реално или в почти реално време. Например, една съвременна система за наблюдение на даден район на град обикновено използва няколко стотици (в Лондон даже хиляди) камери. Ако се търси открадната червена кола от дадена марка, постъпващата информация трябва да бъде обработена максимално бързо за да има ефект търсенето. Често пъти видеокамери се използват за оптимизация на трафик потока (режима на работа на светофарите). В тези случаи решение трябва да бъде предложено в рамките на няколко минути. Ако обаче е необходимо осигуряване на зелена вълна на движение на специално транспортно средство (пожарна, линейка), закъснение от няколко секунди може да бъде фатално.

В настоящата публикация ще бъде разгледан в детайли само един аспект от обработката на видеоинформация, касаещ автоматичното разпознаване на човешки лица в системи за достъп. Въпреки че разглежданият подход е конкретен, много от използваните методи са общи и за останалите задачи, възникващи при обработката на видеопотоци.

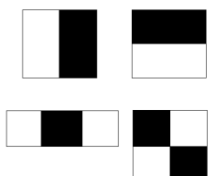
Автоматично разпознаване на човешки лица в системи за достъп

В последните двадесет години проблемът за автоматично разпознаване на човешки лица концентрира вниманието на научната общност, ангажирана в сферата на обработката на изображения и компютърното зрение. Първоначалният ентузиазъм бързо беше сменен с трезва оценка за трудностите, които съпътстват решаването на такава задача. Много автори са единомислени, че автоматичното разпознаване на човешки лица е едно от най-големите предизвикателства в посочената сфера. Множеството алгоритми, които се срещат в научната

литература по въпроса, решават разглеждания проблем с различен успех в различните области на приложение. В най-добрите работи се посочва достигната вероятност за разпознаване надхвърляща 0.95—0.96. В много практически приложения тази вероятност може да е задоволителна. Успехът на разпознаването зависи до голяма степен от условията, при които е снето изображението, подлежащо на сравняване и разпознаване.

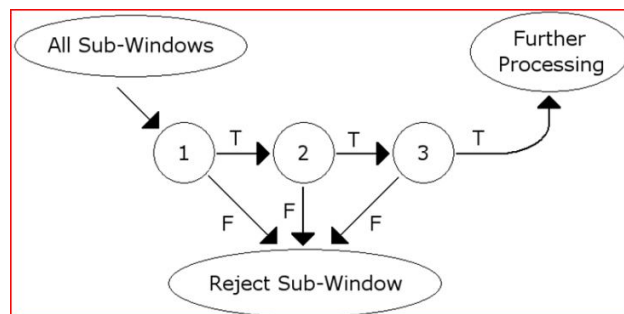
Съществува единодушно мнение на множество учени, ангажирани с този проблем, че наличните алгоритми за разпознаване са съвсем близо до практическо приложение в системите за достъп. Такива системи задължително съдържат подробна видеобаза с изображения на лицата, допускани в обекта (учреждение, военна зона или обект от сферата на сигурността и др.). Предполага се, че мястото за пропускане е подходящо осветено и е снабдено с две или три камери. Камерите заснемат приближаващия индивид и поредицата от кадри в реално време се подлага на обработка. Обработката се извършва от програми за сваляне на отделни кадри от видеопотока, откриване на лице, сегментация на лицето, филтриране и преформатиране и разпознаване. Едно от основните изисквания за успешно разпознаване е гарантирането на някаква минимална разрешаваща способност на изображението на лицето. Това се осъществява с внимателен подбор на оптиката на камерите и тяхното разположение. В реализираната система бе удовлетворено изискването за 100 пиксела между зениците на очите, което резултира в минимален размер на лицето от 160x200 пиксела (общ брой от 32000 пиксела).

Суровите изображения постъпват в някой от стандартизираните формати за изходна информация на цифровите камери: AVI, MPEG4, MJPEG, JPEG, BMP. Първите 3 формата се отнасят за потоци от видеоизображения, а последните два са за режим на сензора "snapshot" при установяване на движение. Аналоговите камери предават аналогова информация, която в VCR (Video Camera Recorder – записващи устройства) устройствата се дискретизира и преобразува в някой от споменатите по-горе формати. Следва предварителна обработка на изображенията, включваща отстраняване на характерния за видеосензорите равномерно разпределен точков шум, и трансформиране на цветното изображение в формати на запис, в които ще се извършва последващата обработка. Обикновено камерите предоставят изображенията във формат RGB или YCbCr. От тези формати полученото изображение се трансформира в монохромен формат. Следва нелинейна обработка с цел промяна на хистограмата на интензивностите на изображението. Основната цел е преодоляване на трудностите при обработка на изображения, възникващи в резултат на неравномерното осветяване на обектите и възникването на сенки. С помощта на тази обработка се усилват елементите, носещи семантиката на изображението и се отслабват тези, които не се използват в последващите етапи на обработка. Най-важната и трудна за реализация задача от предварителната обработка на информацията е откриването и сегментирането на лицето. От многобройните подходи за откриване на лица е избран алгоритъм на Виола-Джонс [1]. Този метод е подобен на корелационния подход, но работи в друго пространство – пространството на Хаар разложението на изследваното изображение. Типично използваните шаблони за откриване на лице са показани на фиг.1. Приложен е каскаден алгоритъм за максимално бързо отсяване на областите, потенциално съдържащи лице (фиг.2). Основни аргументи в полза на направения избор е надеждността и възможността за паралелна реализация.



Фиг.1 Признаци на Виола-Джонс

Резултат от предварителната обработка е изрязано лице/лица при наличие на такова/такива в кадъра.



Фиг.2 Каскаден алгоритъм (T-true, F-false)

За решаване задачата за разпознаване е избран класическият алгоритъм *Principal Component Analysis* (PCA) или *Eigenvalue method for face recognition* [2,3,4]. В много работи, посветени на този проблем, се подчертава, че при стандартни условия на снемане на изображенията този алгоритъм дава резултати, близки до най-добрите. Като добавим и предимството, свързано с неговата простота на реализация и скорост на изпълнение, този алгоритъм беше предпочетен пред много други, появили се на бял свят след него. Причината е в това, че при организиране на максимално подходящи условия за снемане на изображения, този метод работи достатъчно надеждно. Разпознаването се извършва в следните няколко стъпки:

- трансформира постъпилото изображение във вектор с 32000 компонента;
- извежда от получения вектор средното лице, пресметнато а priori, получавайки вектор-девиация на входното изображение;
- проектира получения вектор-девиация върху собствените вектори, получавайки M на брой проекционни коефициента;
- пресмята критерия за близост между пресметнатите проекционни коефициенти на входното изображение с аналогичните групи коефициенти на лицата във видеобазата;
- сортира и сравнява получените резултати, като извежда решението за допускане или не в зоната с ограничен достъп.

Изборът на критерий за близост (разстояние, метрика) играе ключова роля за успешна работа. Образно казано, смяната на метриката в пространството преподрежда елементите в него. Един от важните показатели за избор на дефиницията за разстояние (или критерий за близост, което е едно и също) е свързан с двата типа грешки, които е в състояние да извърши един алгоритъм за разпознаване при конкретното приложение: да се пропусне човек, който няма допуск или да се спре човек, който има допуск. Въпросът опира до две различни множества от разстояния: **а)** множеството от всички най-големи стойности на разстоянието между всеки индивид с достъп до изображенията в базата, които му съответстват; **б)** множеството от най-малки стойности на поредица индивиди без достъп до изображенията в базата. Трябва да се намери такава метрика, при която тези две множества да са отчетливо разделени. Изборът на критерия за сравнение (или за близост) беше извършен на базата на интензивни експерименти със седем различни дефиниции за разстояние. Като най-ефективни бяха избрани две от тях: Yambor angle и Correlation distance. По-долу са приведени няколко стойности от дефинираните по-горе две множества – най-големите от множеството (**а**) и най-малките от множеството (**б**), подбрани от един от провежданите експерименти:

- за Correlation distance: {0.166; 0.174; 0.189} и {0.201; 0.220; 0.225}

- за Yambor angle: {-0.0098;-0.0095;-0.0086;-0.0085;-0.0083} и {-0.0068;-0.0067;-0.0065;-0.0065}

И за двете метрики се вижда, че най-голямото (а това означава и “най-лошото”) разстояние между индивид с достъп до някоя от съответстващите му изображения е все пак по-малко от най-малкото (а това означава и опасно близко) разстояние между случаен индивид и най-близкото до него изображение от видеобазата. Това дава възможност да бъде дефиниран праг, отделящ двете множества.

За успешното разпознаване е необходимо предварително изграждане на видеобаза и обучение на алгоритъма за откриване на лица (*off-line* режим на работа). За избор на най-добрите признаци и обучение е използван един вариант на AdaBoost обучаващ алгоритъм. Изграждането на видеобаза е свързано с чисто организационни дейности, които трябва да бъдат добре планирани и извършени в два етапа. *Първият етап* включва заснемането на определен брой кадри с изображения на всички служители в звеното (учреждението), които ще бъдат обслужвани от системата за допуск. Има определени изисквания към камерите, осветеността и фона, които трябва да бъдат изпълнени. Всяко от изображенията на даден служител трябва да включва определена посока на гледане: директно в камерата, леко вляво, леко вдясно, поглед над камерата, поглед под камерата, изображение с лека усмивка, което не е изключено в такива ситуации (тук изключваме случая, когато служителят прави някакви особени гримаси). Ако даденият индивид постоянно носи очила, то на повечето снимки трябва да е с тях, но трябва да се заснемат и изображения без очила. *Вторият етап* от изграждането на видеобазата с изображения на лица е тяхната обработка, включваща филтрация, нормировка по осветеност, центриране на изображението, и свеждането на броя пиксели до предварително избран стандарт. Нека допуснем, че броят на лицата, които ще имат допуск е N , а снетите от тях изображения са M . Очевидно $M > N$, понеже всеки индивид е заснет няколко пъти. Въвеждат се последователно

всички файлове с изображения. Всяко изображение се трансформира във вектор и се записва в определена матрица. След извършването на всички последващи действия с получените вектори се достига до симетрична матрица с размери ($M \times M$), съдържаща информацията на девиациите на всяко лице от едно средно лице. Именно с тази матрица решаваме задачата за собствените стойности и собствените вектори като получаваме M собствени стойности и M собствени вектора. След подходяща трансформация тези вектори се превръщат в M единични и взаимно перпендикулярни собствени вектори с 32000 компонента. Това е подпространството на собствените вектори, върху което ще бъдат проектирани изображенията на индивидите, които имат допуск до дадената зона.

Паралелни програмни модели

Има различни видове програмни модели. Три от тях са намерили най-широко разпространение: независими (непресичащи се), задачно-ориентирани и нишково-ориентирани. Първите непресичащи се модели са приложими за паралелни приложения, изпълнявани от различни процесори в различни адресни пространства на паметта. В тези случаи се използва MPI (Message Passing Interface – синхронизация чрез съобщения) механизъм за синхронизация на процесорите и получаване на общо решение.

При задачно-ориентираните модели програмистът разделя проблема на множество не/зависими задачи и възлага на диспечера тяхното изпълнение. Диспечерът на задачи има грижата да синхронизира изпълнението на отделните задачи съобразно тяхната взаимозависимост и да оптимизира времето за тяхното изпълнение. Реализацията на този вид модел може да бъде както на универсални процесори (CPU – Central Processing Unit) така и на графични процесори (Graphics Processing Units). Една реализация на задачно-ориентираните модели, която намира особено широко приложение в някои от най-високопроизводителните суперкомпютри, е CUDA (Compute Unified Device Architecture), дефинирана от наложилата се на пазара фирма за графични ускорители NVIDIA. При него програмистът заявява колко копия от една задача (кърнел) да бъдат изпълнени едновременно. Графичните процесори обикновено разполагат с достатъчно ресурс за едновременно изпълнение на повече от 10000 задачи.

При нишково ориентираните модели няколко процесора работят в общо адресно пространство. Работата в общо адресно пространство има много преимущества при предаване на данни от един на друг процесор. От друга страна, за да бъде предотвратена повредата на данните е създаден сложен механизъм за синхронизация и заключване/защита на данните.

Всеки от тези модели е оптимален за решаване на различни специфични задачи. Някои от суперкомпютрите позволяват бързо пренастройване и работа в повече от един програмен модел. За решаване на задачата за разпознаване на лица бе използван непресичащ се програмен модел с MPI.

Паралелна реализация и анализ на скалируемостта

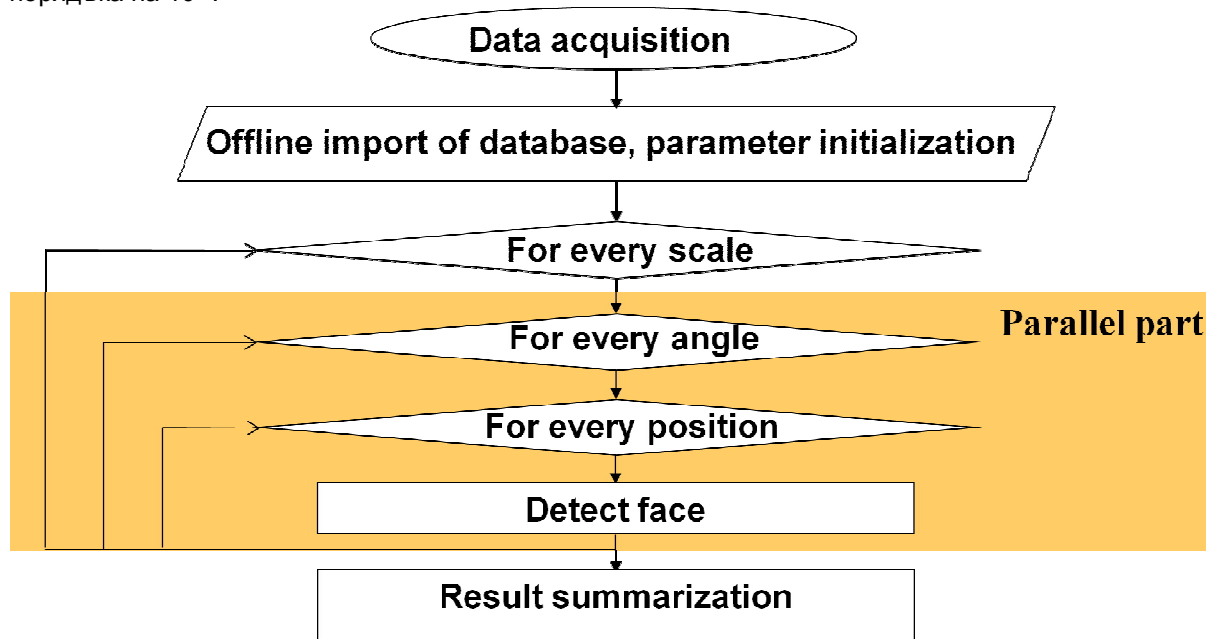
Анализът на възможностите за паралелна реализация на разпознаването на лица показва, че най-подходяща за паралелно изпълнение е откриването на лица (таблица 1).

Таблица 1

| № | Стъпка | Много камери | Много кадри | Мащабно независима | Ротационно независима | Позиционно независима |
|----|--|--------------|-------------|--------------------|-----------------------|-----------------------|
| 1. | Извличане на кадри от видеопоток | да | да | не | не | не |
| 2. | Филтрация на фрейм (anti-noise, blur reduction, contrast enhancement, др.) | да | да | не | не | не |
| 3. | Откриване на лица във фрейм | да | да | да | да | да |
| 4. | Изрязване на откритите лица | да | да | не | не | не |
| 5. | Нормализация на изрязаните изображения | да | да | не | не | не |
| 6. | Изчисляване на метрики/разстояния | да | да | не | не | не |
| 7. | Класификация на получените изображения в съществуващата база данни | да | да | не | не | не |

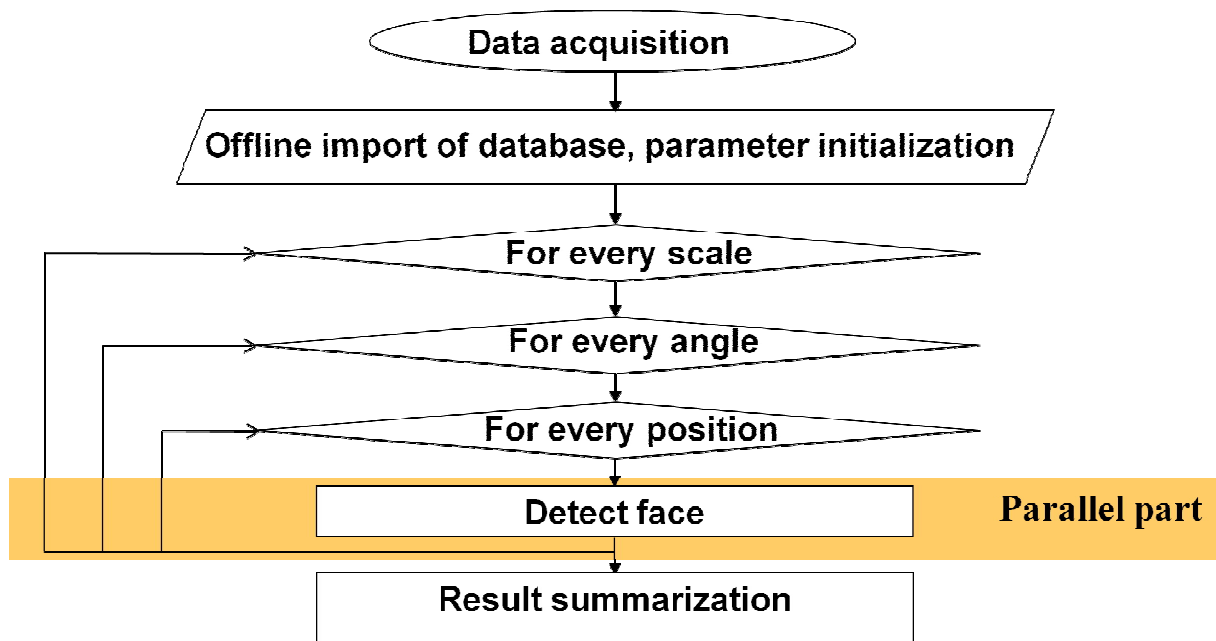
Паралелната програма за разпознаване на лица е написана на език "C" и е инсталирана на суперкомпютър BlueGene/P. Програмата използва функциите на библиотеката MPI (Message Passing Interface). Изображенията на отделните индивиди се съхраняват в pgm- формат, при който няма загуба на информация от компресия.

Трудоемката част на програмата се заключава в проверката за наличие на лице. Тази част е вложена в два цикъла – за различни мащаби на изображението и за различни позиции на лицето в рамките на проверяваното изображение. Това определя и двете възможни реализации на паралелна машина. Първата реализация (фиг.3) се състои в обработката на изображения с различен мащаб на различни процесори. Всичките процесори се зареждат с еднакво количество информация – входното изображение и масивите от коефициенти, които се използват при обработката. Изходната информация е ограничена до няколко числа – при какъв мащаб и къде в изображението (индексите по вертикала и хоризонтала) и намерено лице. За кадри с размер 200x160, размер на шаблона 19x19 и стъпка на изменение на мащаба 1.1 се получават 25 различни мащаба на изображението (т.е. за изображение с малък размер ще се използват 25 процесора). За кадри с мегапикселов размер броят на използваните процесори достига 40 и повече процесора. Основен недостатък на този подход и различното натоварване на процесорите. Отношението в броя обработки на най-натоварения и най-ненатоварения процесори при търсене на лице във всяка възможна позиция (най-тежкия за изчисление случай) в примера с малкото изображение достига 25000 пъти. За мегапикселов размер на изображенията това отношение е от порядъка на 10^6 .



Фиг.3. Паралелна реализация с еднакви входни данни

Втората възможна паралелна реализация на търсенето на лице с помощта на Виола-Джонс алгоритъма е изпълнението и на вложения цикъл по позиция на различни процесори (фиг.4). Тази реализация привлича вниманието с равномерното натоварване на процесорите. Входната информация за процесорите се състои от масивите с коефициенти за обработка и фрагмент от изображението, в който се търси лице. Отделянето на фрагменти увеличава до известна степен количеството изчисления в последователната част на програмата. За търсене на лице при така цитираните стойности на параметрите (за малко изображение 200x160) броят на необходимите процесори е 123550.



Фиг.4. Паралелна реализация с различни входни данни

Алгоритъмът за класификация/разпознаване има смисъл да бъде реализиран на паралелна архитектура ако базата съдържа 1500 и повече изображения. Такава задача не е нереалистична, като си припомним, че от всеки индивид, на който му е предоставен достъп, се снемат по няколко изображения (5, 6 или 7). Това означава, че за предприятия с личен състав в размер на 300—400 души е необходима паралелна реализация на алгоритъма.

На фиг. 5 и фиг. 6 са показани някои резултати от работата на алгоритмите за разпознаване на лица.



Фиг. 5 Откриване на лица: а) две лица в различен мащаб, множество намерени; б) краен резултат



Фиг. 6 Разпознаване на лице: а) включено в базата; б) което не е включено в базата

Заклучение

Решаването на много от проблемите на съвременното общество налага обработката на огромно количество видеоинформация. Модерните персонални компютри нямат необходимата мощност да се справят с тази задача. Използването на суперкомпютърни архитектури за целта, макар и да е голямо предизвикателство, е наложително. Досега суперкомпютрите се използват

преди всичко за работа с огромни модели, докато разглежданата задача се отличава основно с дълбочинна обработка на огромно количество входни данни. В статията е разгледано приложението на Blue Gene/P за откриване на лица. Използван е независим програмен модел с използване на MPI. Интерес би представлявало използването и на други програмни модели и сравняване на техните характеристики при решаване на поставената задача, което е обект на бъдещи изследвания.

Благодарност

Част от резултатите, публикувани в тази статия, са получени в рамките на договор „Проектиране на софтуер за събиране и йерархична обработка на информацията от стотици сензори” и посредством подкрепата на Националния център за суперкомпютърни приложения, осигурил достъпа и използването на суперкомпютъра IBM Blue Gene/P.

Литература

- [1] Paul Viola, Michael Jones, “Robust Real-time Object Detection”, Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing, and Sampling.
- [2] M. Turk and A. Pentland, “Eigenfaces for Recognition”, Journal of Cognitive Neuroscience, vol. 3, No 1, 1991.
- [3] Javier Ruiz-del-Solar, P. Navarrete, “Eigenspace-based Face Recognition: A comparative study of different approaches”, IEEE Transactions on Systems, Man & Cybernetics, Part C., Vol. 16, No. 7, 817-830, 2002.
- [4] B. A. Draper, K. Baek, M. S. Bartlett, J. R. Beveridge, “Recognizing faces with PCA and ICA”, Computer Vision and Image Understanding, vol. 91 , Issue 1-2, July, 2003.