

# Compressió de microarrays d'ADN

Iván Pueyo Navarro

**Resum**— La tecnologia relacionada amb la creació d'imatges de microarray d'ADN és una eina de gran importància en el descobriment de l'estructura i funcionament de la nostra informació genètica. Dins d'aquest camp, la detecció del comportament de determinats gens sota condicions específiques adquireix una gran rellevància. La grandària de les imatges de microarray sol ser de mitjana elevada, a causa de la gran quantitat de gens que s'analitzen i al fet que s'intenta mantenir en les imatges el major contingut d'informació possible. Aquestes són generades en grups de dues imatges i en escala de grisos. La finalitat d'aquest projecte és facilitar l'anàlisi de les imatges de microarray als especialistes a través de la combinació del parell d'imatges per formar una imatge en color en RGB, i alhora reduir la grandària d'aquestes imatges per no desapropiar l'espai físic. En el camp dels microarrays d'ADN perdre informació en les imatges equival a perdre dades d'anàlisi, per aquesta raó la reducció de la mida de les imatges és realitzada a través de la compressió sense pèrdua amb l'estàndard JPEG 2000, el que permet reduir la grandària d'aquestes alhora que manté la informació original continguda en elles.

**Paraules clau**— Microarray ADN, compressió JPEG 2000 microarray, imatge microarray ADN RGB, microarray color.

**Abstract**— The technology related to the DNA microarray image generation is a very important tool in the discovery of the structure and operation of our genetic information. Within this field, the detection of the behavior of some certain genes has a great relevance. The size of the images of microarray is at average high due to the large number of genes analyzed and the fact that it attempts to maintain the most amount of information in the images. These are generated in groups of two in grayscale. In order to facilitate the analysis of microarray images to specialists it is proposed to combine the pair of images to form a color image in RGB, and at the same time reducing the size of images to avoiding to squander the physical space. In the field of DNA microarray, losing information in the images is equivalent to losing data analysis, therefore reducing the size of images is done through lossless compression with the JPEG 2000 standard, which allows to reduce their size while maintaining the original information contained in them.

**Index Terms**—DNA microarray, JPEG 2000 microarray compression, DNA microarray RGB imatge, color microarray.



## 1 INTRODUCCIÓ

ELS microarrays d'ADN [1], [2], [3], [4], [5] són una tecnologia molt utilitzada en l'àmbit de la biologia i biomedicina [6], [7], [8], [9], des de la investigació del càncer fins al control de plagues, mitjançant l'anàlisi concurrent d'un gran nombre de gens. El funcionament comença amb marcar dues mostres orgàniques, la de control i la que es vol estudiar, amb dos colorants fluorescents diferents: Cy3 -verd- i Cy5 -vermell-. Un dispositiu robòtic disposa de forma ordenada i matricial els gens presents en els fragments d'ADN dels teixits, en un portaobjectes, on cadascuna de les localitzacions d'aquest gens s'anomena spot. Seguidament, s'utilitza un escàner el qual aplica una font de llum vermella i una altra verda per poder mesurar la intensitat lumínica de cada spot, donant com a resultat dues imatges en escala de

grisos, on les intensitats més properes al blanc denoten una major activitat genètica. Es comparen ambdues imatges, focalitzant-se en la diferència d'intensitat en cadascuna respecte cada spot. Aquesta informació resulta d'utilitat a l'hora de determinar la funcionalitat d'un gen en diferents condicions. En la Figura 1 es pot veure la representació esquemàtica del procés anteriorment descrit.

Les imatges en escala de grisos generades poden excedir amb freqüència altes resolucions com 4000 x 13800 píxels. Degut principalment al gran rang d'intensitats que pot presentar un spot, és necessària la utilització de 16 bits per emmagatzemar cada píxel de la imatge. Tant és així, que depenent de la resolució de sortida de cada imatge, la seva mida pot ser de més de 230 Megabytes.

Aquestes dues imatges de microarray d'ADN en escala de grisos també poden combinar-se per obtenir una única imatge en format de color RGB [10]. L'objectiu és facilitar

- E-mail de contacte: [ivan.pueyo@e-campus.uab.cat](mailto:ivan.pueyo@e-campus.uab.cat)
- Menció realitzada: *Tecnologies de la Informació*.
- Treball tutoritzat per: Joan Serra Sagristà (Departament d'Enginyeria de la Informació i de les Comunicacions)
- Curs 2014/15

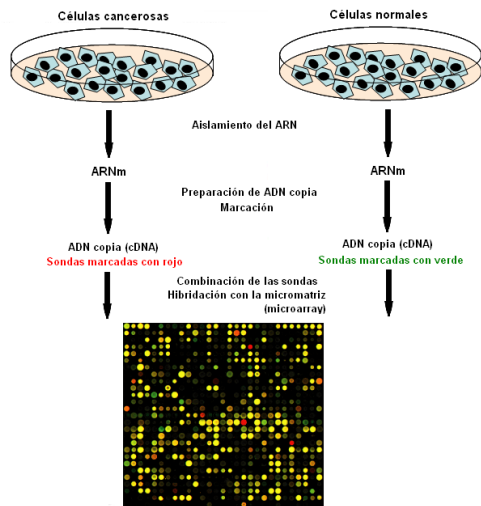


Figura 1. Esquema del principis dels microarrays d'ADN [11].

la identificació visual de l'expressió dels gens en cada mostra, tal i com es pot veure en la Figura 2, on el spot verd indica que el gen s'expressa més en la mostra de control, vermell si el gen s'expressa amb més intensitat en la mostra a estudiar, negre si no hi ha expressió del gen en cap de les dues mostres, groc si es troba igualment expressat en ambdues mostres, i variacions de groc – des de tonalitats més properes al verd fins més properes al vermell – que mostren diferents variacions en l'expressió del gens entre les dues mostres.

El model de color RGB està compost per tres canals amb els seus respectius colors de llum primaris: vermell, verd i blau. D'aquests tres canals de color, la imatge en escala de grisos amb el colorant fluorescent Cy3 forma el canal de color verd, la imatge en escala de grisos amb el colorant fluorescent Cy5 forma el canal de color vermell, i el canal blau té valor zero al no tenir cap imatge que aportí informació pel color blau. Amb aquest procés es forma una imatge en color RGB a partir de les dues imatges en escala de grisos, i encara que al llibre [10] s'explica el procés per una imatge en color RGB de 24 bits a partir de dues imatges en escala de grisos de 16 bits comprimides a 8 bits cadascuna, també és possible el mateix procés amb una imatge final en color RGB de 48 bits per evitar comprimir les dues imatges originals.

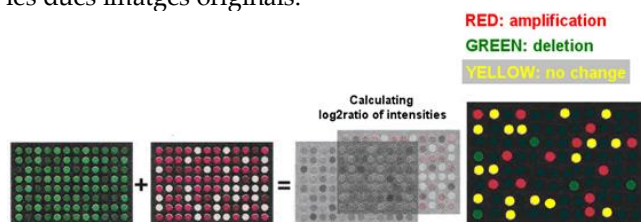


Figura 2. Imatge modificada de la combinació d'imatges de microarrays Cy3 i Cy5 en una imatge en color RGB [9]

L'estudi tant de les imatges en escala de grisos com les imatges en color RGB requereix el seu emmagatzemament de forma indefinida i sense pèrdua d'informació, per tant, és necessari introduir un mètode de compressió eficient sense pèrdua que asseguri la integritat de les dades. Les

imatges de microarrays són utilitzades amb finalitats mèdiques, això implica que s'han d'adherir a l'estàndard d'imatge digital i comunicació en medicina (DICOM) [12] perquè puguin ser utilitzades en els centres de salut. Per aquesta raó, s'ha elegit per a aquest treball un dels millors estàndards de compressió d'imatges acceptats per la DICOM i amb suport d'imatges de profunditat de color de 16 bits: el mètode de compressió sense pèrdua JPEG 2000 [13].

## 1.1 Objectius

L'objectiu principal d'aquest projecte és: **Desenvolupar una aplicació per la compressió d'una imatge en model de color RGB d'un microarray d'ADN, mitjançant l'estàndard de compressió JPEG 2000 sense pèrdua.** Aquest objectiu està format per un subconjunt d'objectius, els quals en conjunt formen l'objectiu final. A continuació es troben els quatre objectius en ordre i en format de llista:

1. Una millor comprensió de què és un microarray d'ADN, a través de l'anàlisi del procés que comença amb la selecció de dues mostres genètiques, i finalitza amb l'obtenció de dues imatges en escala de grisos.
2. La generació d'una imatge en color (RGB) a partir de les dues imatges en escala de grisos que s'obtenen d'un microarray d'ADN.
3. La recerca del funcionament de l'estàndard de compressió JPEG2000 sense pèrdua i com es poden adaptar les fases més importants al tipus d'imatge en color, que es genera a partir de dues imatges en escala de grisos de microarrays d'ADN. Aquesta compressió és necessària per poder emmagatzemar les imatges per posteriors anàlisis genètics, al ser imatges amb una mida gran i evitant malgastar espai.
4. Trobar la transformació espectral més adequada, tant per generar una adequada compressió sense errors com per evitar la pèrdua de qualitat. Les transformacions candidates són: 5/3 DWT, POT, DPCM i R-Haar.

## 1.2 Estat de l'art

L'any 2003, R. Jornsten juntament amb W. Wang, B. Yu i K. Ramchandran van proposar el LOCO segmentat (SLOCO) [5], [14]. Hi ha tres aspectes a tenir en compte [14]. Els spots i el background són codificats separatament. S'utilitza un "UQ-adjust quantizer" que millora la quantització de la predicció dels errors. Té en compte la variació màxima dels límits d'error de píxel.

En 2004, l'equip integrat per J. Hua, Z. Liu, Z. Xiong, Q. Wu i K. Castleman va publicar el software BASICA [5], [15]. Aquest processa les imatges de microarray per eliminar les irregularitats en la segmentació i utilitza un algoritme de truncament optimitzat anomenat EBCOT, el qual aconsegueix una taxa de distorsió òptima en la codificació amb pèrdua mentre encara manté un gran rendiment en la compressió sense pèrdua. En el mateix any 2004, S. Lonardi i Y. Luo van presentar l'algoritme

MicroZip [1], [5]. Després de la segmentació de la imatge en spot i background, aquest utilitza un mecanisme amb pèrdua per codificar el byte menys significatiu dels píxels de background i la resta amb un mecanisme sense pèrdua. En 2006, Y. Zhang, R. Parthe i D. Adjeroh [2] van proposar un mètode sense pèrdua utilitzant també segmentació, en el qual una vegada dividida la imatge, s'utilitza un esquema predictiu per als bits més significatius de cada píxel i una predicció per correspondència aproximada parcial (PPAM) per als bytes menys significatius.

A partir de les proves realitzades al document [5], s'exposen tres mètodes o tècniques de compressió: Compressió multicomponent, palette reindexing i Histogram swap transform (HST).

La compressió multicomponent és el mètode de compressió més efectiu en el que respecta a la compressió conjunta del parell d'imatges de microarray en escala de grisos. Aquest aprofita la similitud entre les imatges, a l'aparèixer en la gran majoria de casos els spots en posicions i intensitats coincidents, per codificar aquesta redundància en un sol fitxer.

La palette reindexing ha demostrat ser un mètode eficaç per millorar el rendiment de l'estàndard JPEG 2000 en els diversos conjunt d'imatges de microarrays d'ADN amb els quals s'ha experimentat. Tenint en compte que les imatges a color poden ser representades utilitzant una matriu índex i una paleta de color - associa la matriu índex amb els colors originals -, i que en les imatges en escala de grisos la paleta de colors passaria a identificar intensitats en lloc de colors; el palette reindexing reordena aquesta paleta d'intensitats mitjançant una heurística que consisteix en minimitzar l'entropia de primer ordre de les diferències dels consecutius índexs.

Finalment, el HST és capaç de produir millores de rendiments molt similars als dos anteriors en JPEG 2000. Es basa en el fet que les imatges de microarray d'ADN presenten una distribució molt asimètrica de la distribució d'intensitats. La majoria dels píxels dels conjunts d'imatges de microarray d'ADN utilitzats al document [5] tenen baixa intensitat, mentre que valors majors són menys freqüents. Aquest fet és utilitzat per l'HST amb la finalitat d'obtenir histogrames tan simètrics com sigui possible.

Per finalitzar és necessari parlar dels algorismes Battiato [5], [16] i Neves [5], [17], els quals proporcionen el millor i segon millor rendiment, respectivament. En 2006, Neves i Pinho van publicar un mètode que codifica els bitplanes - conjunt de bits corresponents a una posició de bit donada en cada valor binari que representa el senyal - per ordre d'importància des de l'últim -menys significatiu- fins al primer, el qual és el més significatiu i per tant no pot ser comprimit. En aquest s'utilitzen models de context 3D, els quals reben aquest nom en codificar un bit compost de bits provinents dels anteriors bitplanes. La forma d'utilitzar els píxels en el context - plantilla de context- no canvia i es troba prèviament calculada per un procediment de prova i

error que genera una millora del rendiment mitjà de la compressió. En 2009 [18], van ampliar el seu mètode perquè la plantilla de context fos feta especialment per cada imatge, la qual cosa també origina un augment en el cost de còmput. En 2009, Battiato i Rundo van presentar un algorisme que utilitza les xarxes neuronals cel·lulars per segmentar les imatges en foreground - spot - i en background. La imatge del foreground és comprimida utilitzant l'estàndard de compressió sense pèrdua PNG. Es transformen els 16 bits del background en dues imatges indexades de 8 bits cadascuna utilitzant un algorisme personalitzat per reduir la grandària de la paleta de colors - intensitats en cas d'imatges en escala de grisos -. La paleta d'aquestes imatges es reordena utilitzant un algorisme que minimitza l'entropia de les diferències locals, i les matrius índex restants també es codifiquen amb PNG.

### 1.3 Metodologia

Una metodologia de desenvolupament de software [19], [20], [21] es refereix a una filosofia o forma de treballar específica, utilitzada per estructurar, planificar i tenir sota control el procés de desenvolupament d'un projecte. Gran part de la probabilitat d'èxit o fracàs d'un projecte recau en la correcta elecció de la metodologia necessària per la seva consecució en el període de temps fixat.

Els models tradicionals més coneguts i utilitzats són 4: desenvolupament incremental, desenvolupament iteratiu, model en espiral i model seqüencial o en cascada.

Analitzant el quatre models, s'arriba a la conclusió que hi ha models que no encaixen amb els objectius del projecte. El model de desenvolupament incremental té com a objectiu reduir el temps de desenvolupament dividint el projecte en intervals incrementals superposats, però també requereix que els increments siguin funcionalment independents i aquest no és el cas dels objectius i tasques que deriven del projecte. El desenvolupament iteratiu es centra en capturar millor els requisits canviants, es trenca el projecte en iteracions de diferents longituds, cadascuna produint un producte complet i entregable; el problema esdevé en que el projecte no es pot entregar en diferents entregables resultat de les diferents iteracions. Un model en espiral té en compte les característiques del cicle de vida bàsic combinat amb el prototipatge, el qual es basa en el desenvolupament iteratiu, per tant es troba amb el mateix problema d'abans. Per últim, es troba el model seqüencial o en cascada, més conegut per Waterfall. És la metodologia que s'utilitzarà en el desenvolupament d'aquest projecte. S'inicia amb un anàlisi complet dels requisits o objectius del projecte, es continua amb la implementació del disseny i finalment, el sistema es testja i entrega al client.

El model Waterfall divideix el sistema en etapes seqüencials inamovibles no repetibles, codependents i amb una revisió final per cada etapa. És per tant, necessari establir primerament el conjunt d'etapes del sistema de forma correcta, ja que aquestes no es podran moure endavant ni endarrere; i tenir en compte que per passar

d'una etapa anterior a una altra posterior, és requisit indispensable que aquesta etapa anterior hagi finalitzat correctament i superat la seva revisió final. A continuació s'exposen les quatre fases principals del model Waterfall orientades a aquest projecte:

- 1 **Anàlisis.** En aquesta etapa s'analitzen quins objectius es vol cobrir en el projecte a través d'una investigació inicial sobre els microarrays d'ADN i l'estàndard de compressió d'imatges JPEG 2000, la planificació de les fases/activitats del projecte, i la planificació temporal. Aquesta fase engloba les fases Inici i Planificació del projecte.
- 2 **Disseny.** És descompon el projecte en dos mòduls, generació d'imatge en color RGB i compressió d'imatge en color RGB, amb els seus propis tests personalitzats. En aquesta fase es genera un disseny preliminar i detallat, on es defineix l'estructura que tindrà l'aplicació final i els algorismes i llibreries que s'utilitzaran en la implementació. Aquesta fase correspon a la fase amb el mateix nom del projecte.
- 3 **Codificació.** És l'etapa on s'implementa el codi font de l'aplicació, fent ús de proves i tests de correcció d'errors dissenyats en la fase de Disseny per la generació de la imatge en color i la compressió de la imatge en color. En el projecte es subdivideix en Desenvolupament i Test i Proves.
- 4 **Validació.** És l'última etapa del projecte on s'executa l'aplicació composta pels mòduls dissenyats, codificats i testejats en fases anteriors. També es generen els documents relacionats amb la finalització del projecte com l'especificació tècnica del projecte, el dossier del TFG i aquest article.

## 2 DESENVOLUPAMENT

El desenvolupament del projecte s'ha dividit en dues parts: el mòdul per generar imatges en color RGB i el mòdul per a la compressió, validació i visualització d'imatges en color RGB.

### 2.1 Mòdul generador d'imatges en color RGB

Aquest mòdul té com a funció poder generar i emmagatzemar una imatge en color RGB a partir de dues imatges de microarray d'ADN de 16 bits i en escala de grisos. Per a la creació d'aquest mòdul s'ha fet ús de l'entorn de desenvolupament integrat (IDE) Eclipse Luna versió 4.4.0, el llenguatge de programació orientat a objectes Java versió 1.8.0, i les llibreries de tractament d'imatges en format PNG, TIFF, RAW: `jai_imageio-core-1.3.0` i `ij.jar`.

El disseny del mòdul segueix el diagrama de flux de la Figura 3. Les dues imatges d'entrada amb fluorescència Cy3 i Cy5 són carregades al programa a través d'una finestra que permet a l'usuari explorar els directoris del sistema fins a la ruta d'aquestes imatges. El programa té dos carregadors, un per a imatges amb extensió JPEG, JPG, PNG, PGM, TIFF o TIF; i un altre exclusiu per a imatges en

format RAW, on prèviament ens sol·licitarà introduir l'ampada i altura de les imatges per poder carregar-les.

Una vegada s'han carregat i validat correctament les imatges es llegeixen píxel per píxel cadascuna i s'emmagatzemen temporalment al programa. Tenint ja les dues imatges llegides i emmagatzemades, es crea una tercera imatge en escala de grises - tots els seus píxels tenen valor zero -, obtenint d'aquesta forma els tres components necessaris per crear una imatge en color RGB -imatge amb canal vermell, verd i blau -. Se li assigna a cada canal la imatge que li correspon, sent el canal verd amb la imatge amb fluorescència Cy3, el canal vermell amb la imatge amb fluorescència Cy5, i el canal blau amb la imatge creada anteriorment - no tenim cap altra imatge amb informació per a aquest canal -. Per finalitzar, la imatge en color RGB generada es guarda en el mateix directori que les imatges d'entrada i amb extensió TIF.

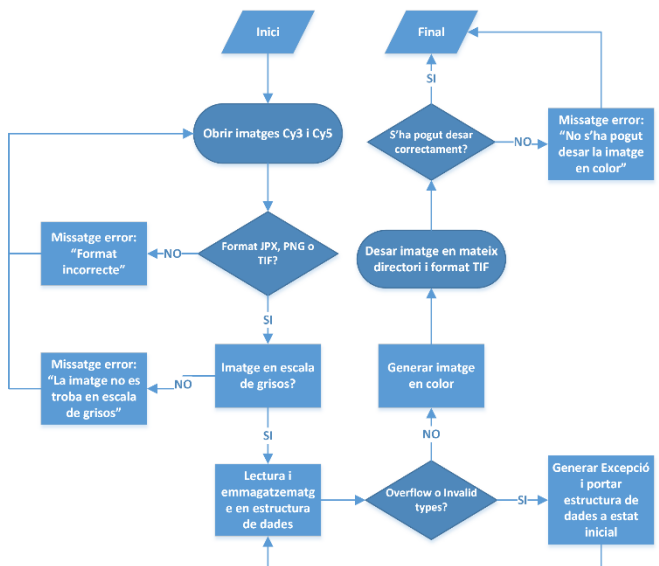


Figura 3. Diagrama de flux del mòdul de generació d'una imatge en color a partir de dues imatges en escala de grisos de microarray.

Aquest mòdul també incorpora un conjunt de funcionalitats més enllà del fet de carregar dues imatges de microarray d'ADN en escala de grisos, combinar-les amb una imatge nul·la per al canal blau, generar la imatge en color RGB amb les tres imatges anteriors, i guardar la imatge en el mateix directori que les imatges d'entrada. Són les següents:

- Es permet la combinació dels canals vermell, verd i blau -R, G i B- mitjançant les dues imatges en escala de grisos de microarray d'ADN. La combinació pot ser la desitjada per l'usuari i no tan sols permet millorar subjectivament la visualització de la imatge en color resultant, sinó que també pot ajudar a l'usuari a compensar la dificultat que pugui tenir en distingir diferents colors, tonalitats de color o compensar un determinat nivell de daltonisme. Entre aquestes combinacions, per exemple, podríem trobar la imatge del canal vermell duplicada i assignada al

canal blau mentre es manté inalterat el canal verd o la imatge del canal verd assignada al canal blau, deixant el canal verd a nul i el canal vermell inalterat. La Figura 4 proporciona un exemple del seu ús.

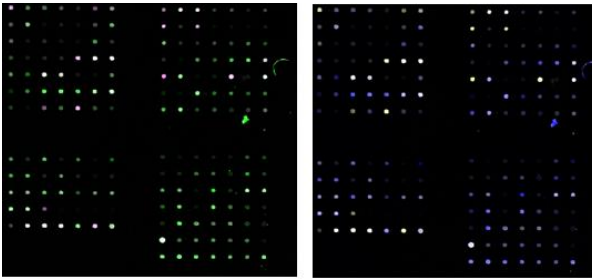


Figura 4. Dues combinacions diferents de les imatges Def663Cy3.raw i Def663Cy5.raw del set isrec\_raw. Canals de la imatge de l'esquerra R, G, B; i R, R, G de la dreta.

- Implementat un comptador del nivell gamma de la imatge que permet ajustar la lluminositat d'aquesta, creant ombres més fosques en augmentar el seu nivell o fent que les zones fosques siguin més clares en reduir-ho. Els valors gamma estan limitats en l'aplicació a un mínim de 0,1 i un màxim de 3, considerant que valors menors al límit inferior fan a la imatge excessivament brillant, i valors majors al límit superior, massa fosca. Resulta també una ajuda a l'usuari en la visualització de les imatges, ja que és necessari compensar certes propietats de la visió humana, amb la finalitat de maximitzar la percepció de la llum i el color. En la Figura 5 es pot apreciar la seva utilització.

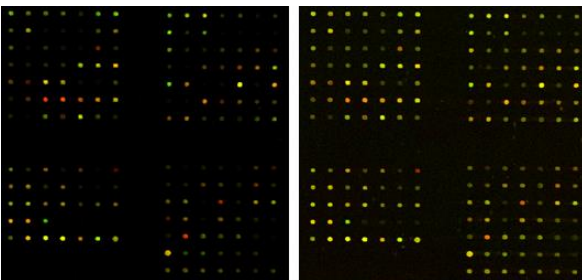


Figura 5. Dues combinacions idèntiques de les imatges Def667Cy3.raw i Def667Cy5.raw del set isrec\_raw. La imatge de l'esquerra manté el seu nivell gamma original -per defecte 1- i la de la dreta té un nivell de 0,5.

- Afegida una opció per duplicar el valor dels píxels que conformen la imatge en color RGB generada a partir de les imatges en escala de grisos. Dins de l'espectre electromagnètic visible per l'ull humà, cada persona percep de forma diferent la intensitat dels colors, depenent en gran manera de la sensibilitat dels seus tres tipus de cons -sensibles a la llum vermella, blava i verda-, per tant això influeix en les tonalitats que pot reconèixer, així com en la seva dificultat. És per aquesta raó, que és probable que no tots els usuaris, encara tenint a la seva disposició la possibilitat de modificar el gamma i la combinació de canals que formen la imatge en color, siguin capaços de distingir

correctament totes les tonalitats de colors presents en aquesta imatge. Per facilitar aquesta tasca, és possible activar aquesta opció, augmentant individualment el valor de cada píxel i fent d'aquesta manera més evident el contrast de colors en la imatge. La Figura 6 mostra el seu efecte.

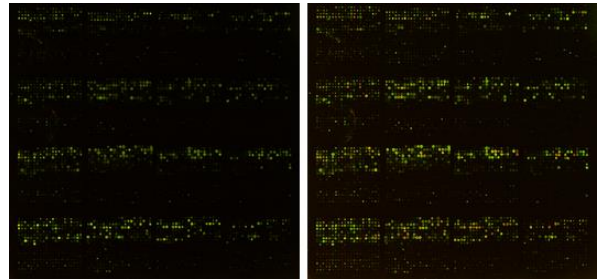


Figura 6. Dues combinacions de les imatges 1230c2G.tif i 1230c2R.tif del set apoa1. En la imatge de l'esquerra es manté el valor original de tots els seus píxels i en la de la dreta s'han multiplicat per 2.

## 2.2 Mòdul compressor, validador i visualitzador d'imatges en color RGB

El mòdul que realitza la compressió, validació i visualització d'imatges en color RGB de microarrays d'ADN s'ha fet independent del mòdul per a la generació d'aquestes imatges. Si bé el mòdul de generació d'imatges en color RGB està realitzat amb el llenguatge de programació orientat a objectes Java en la seva versió 1.8.0, el mòdul de compressió està escrit en el llenguatge de programació Shell scripting Bash, fent ús de l'eina de software per a compressió d'imatges Kakadu 7.4 en la seva versió gratuïta [22] i el sistema operatiu Ubuntu 14.04 x64.

El funcionament del mòdul segueix el diagrama de flux de la Figura 7. Aquest mòdul està format per 4 scripts en total: "MainDNAmicroarrayCompress", "Compression", "kdu\_expand\_vol", "Validation".

El programa comença quan l'usuari executa per consola el script "MainDNAmicroarrayCompress", el qual és el script principal del programa i el que crida als altres scripts, juntament amb la ruta de les dues imatges en escala de grises i les seves dimensions. Pel que respecte a l'extensió de les imatges, en cas que aquestes no es trobin en format RAW, es transformaran a format PGM mitjançant l'eina tiffopnm -normalment instal·lada per defecte en Ubuntu 14.04 x64 - i a continuació a format RAW eliminant la seva capçalera. La raó darrere de la conversió de les imatges a RAW prové d'un problema desconegut amb l'eina Kakadu, la qual encara al comprimir sense mostrar errors imatges amb altres formats, com TIFF, no realitza correctament la combinació de canals i genera com a resultat una imatge comprimida en escala de grisos -només si més d'un component té un format d'imatge diferent a RAW -.

Seguidament, es crida al script "Compression" -genera la imatge del canal blau i realitza la compressió - i fent ús de l'eina ImageJ 1.48 que permet visualitzar imatges RAW

en Ubuntu -també es troba per a altres S.O.-, s'escriu un script per generar una imatge en negre -imatge omplerta amb zeros- amb les mateixes dimensions de les altres dues imatges de forma transparent per a l'usuari.

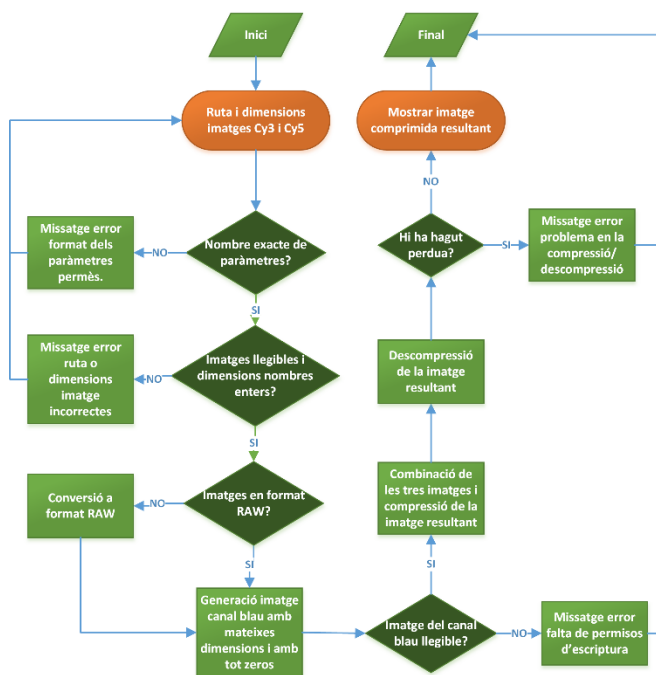


Figura 7. Diagrama de Flux del mòdul de compressió, validació i visualització d'una imatge en color a partir de dues imatges en escala de grisos de microarray d'ADN.

Una vegada generada aquesta imatge per al canal blau, s'executa l'eina Kakadu amb una modificació del mòdul de compressió, a més de canvis i addicions en els seus paràmetres. La compressió s'ha dut a terme utilitzant les dues imatges en escala de grisos introduïdes en iniciar el script, la imatge generada en el pas anterior per al canal blau, les dimensions de les tres imatges, i tenint en compte que cadascuna d'aquestes imatges té 16 bits per píxel. El format de sortida utilitzat és JPX. Format de la família JPEG 2000 que dona accés tant a la primera part de l'especificació de l'estàndard com a la segona, uns altres de la mateixa família com a J2C o JP2 es troben més restringits.

Un aspecte rellevant d'aquesta compressió és la seva realització fent ús de la Transformació Discreta Wavelet (DWT) - transformada matemàtica que representa una senyal en termes de versions traslladades i dilatades d'una ona finita -, en la seva versió de compressió reversible sense pèrdua 5/3 DWT, amb un offset de 4, un rang de 0 a 4 nivells espectrals i de 0 a 5 nivells o fases espacials - a l'apartat 3. Resultats s'indica la implicació i els resultats obtinguts per a cadascun -. Les transformacions divideixen la imatge en un conjunt de sectors amb uns nombres que representen aspectes d'aquesta associada a un rang de freqüències determinades i un àmbit espacial, també anomenats coeficients. La 5/3 DWT utilitza únicament coeficients enters, pel que la sortida no requereix arrodoniment - com succeeix si els coeficients són nombres reals - i no es genera cap soroll - pèrdua de qualitat -.

Altres aspectes rellevants han estat l'ús de la Transformada Multicomponent (MCT), la qual és efectiva reduint les correlacions -si existeixen- entre els múltiples components en una imatge multicomponent, reduint així la redundància i incrementant el ràtio de compressió; i del perfil ICC -s'explica detalladament a continuació- amb la matriu:

$$\begin{pmatrix} 0.6186846 & 0.3313352 & 0 \\ 0.3190093 & 0.5616254 & 0 \\ 0.0290009 & 0.4220004 & 0 \end{pmatrix}$$

Pel que fa al mòdul de compressió de Kakadu, kdu\_compress, s'ha substituït l'original proveït en la versió gratuïta de Kakadu 7.4 per una versió modificada de kdu\_compress en la versió de Kakadu 7.5 per Linux 64 bits, realitzada per Miguel Hernández.

Aquest nou mòdul permet especificar la matriu del perfil ICC. Dins de l'àmbit de la gestió del color aquest perfil és un conjunt de dades que caracteritza a un dispositiu d'entrada o sortida de color, o espai de color, segons els estàndards del Consorci Internacional del Color (ICC). El perfil ICC descriu els atributs de color de la imatge mitjançant un espai de connexió de perfil (PCS) denominat CIE XYZ, model en el qual la Y representa la lluminositat; Z és aproximadament igual a la tonalitat blava, i X és una barreja amb una aproximació a la tonalitat vermella. L'espai de color CIE XYZ és un dels primers espais de color definits i va servir com a base a l'espai de color RGB, no obstant això, a diferència d'aquest últim que es basa en colors primaris additius - vermell, verd i blau -, el CIE XYZ es basa en tres colors primaris imaginàries espectrals - X, Y i Z - que representen el color. Una vegada en aquest espai de color, es pot passar al model sRGB, aquest últim és també conegut com Standard RGB i és un espai de color més petit que el RGB, ideat inicialment per al seu ús en monitors, pantalles de televisor, projectors i un llarg etcètera. El motiu darrere d'aquesta doble transformació és utilitzar aquesta matriu del perfil ICC per poder modificar la visualització final de la imatge comprimida, com per exemple l'associació de quina imatge a quin canal o el valor gamma.

El procés d'utilització del perfil ICC segueix el diagrama de flux de la Figura 8. Realitzada la compressió i generada una imatge JPX, es realitza el procés invers -descompressió-, obtenint els components de la imatge que seran multiplicats de forma matricial per l'esquerra - cal recordar que el producte de matrius no és commutatiu - per una matriu 3x3 - la seva dimensió és estàndard - els valors de la qual modifiquen el gamma i la combinació de canals de la imatge.

Tot seguit, una vegada es tenen els components a l'espai de color CIE XYZ amb l'anterior producte, Kakadu aplicarà la seva matriu 3x3 (1) al producte anterior - multiplicat igualment per l'esquerra -, i obtenint una imatge a visualitzar en sRGB. Per poder modificar correctament els

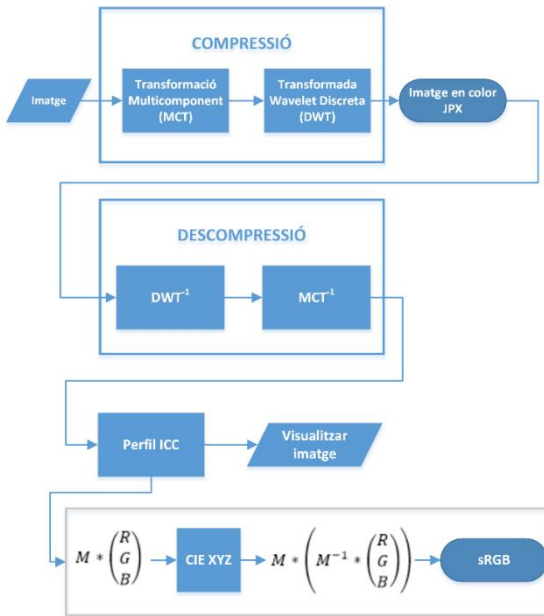


Figura 8. Diagrama de Flux amb el procés del perfil ICC afegit pel mòdul de compressió modificat.

canals i el gamma de la imatge és necessari que la matriu introduïda en el perfil ICC sigui el producte de la matriu inversa  $M^{-1}$  i la matriu desitjada  $A$  -s'explica al apartat 3.Resultats- perquè en aplicar Kakadu el pas de CIE XYZ a sRGB la matriu  $M$  i la seva inversa s'anul·lin, deixant únicament la matriu que es desitja aplicar a la imatge.  $M \times M^{-1} \times A = A$ .

$$\begin{pmatrix} 3.2404542 & -1.5371385 & -0.4985314 \\ -0.9692660 & 1.8760108 & 0.0415560 \\ 0.0556434 & -0.2040259 & 1.0572252 \end{pmatrix} \quad (1)$$

Una vegada completada la compressió i finalitzada la transformació amb el perfil ICC, es procedeix a descomprimir la imatge resultant mitjançant una modificació del script "kdu\_expand\_vol", creat per Ian Blanes, el qual utilitza el mòdul kdu\_expand de Kakadu per obtenir els tres components que formen la imatge en color. Es crida al script "Validation" i fent ús de l'eina de Ubuntu, "diff", es comprova que les imatges corresponents als canals vermell i verd coincideixen amb les mateixes localitzades en les rutes introduïdes a l'inici del script principal, mostrant un missatge per consola indicant si la compressió s'ha realitzat o no correctament. Finalment, el mòdul kdu\_render de Kakadu permet convertir la imatge en color JPX a format TIF i obrir-la amb el visualitzador d'imatges per defecte de Ubuntu -normalment Gimp-.

### 3 RESULTATS

En aquest apartat es presenten els resultats obtinguts en completar l'objectiu principal del projecte: Desenvolupar una aplicació per a la compressió d'una imatge en model de color RGB d'un microarray d'ADN, mitjançant l'estàndard de compressió JPEG 2000 sense pèrdua.

El primer que es presenta és l'aplicació realitzada en

Java per a la generació d'imatges en color RGB de microarrays d'ADN, la qual es mostra un exemple de la visualització de la imatge en color generada en la Figura 9. La descripció d'aquesta aplicació, així com el seu funcionament, ha estat explicada a l'apartat 2.1 Mòdul generador d'imatges en color RGB. L'aportació d'aquesta aplicació al projecte serà explicada a l'apartat 4.Conclusions.

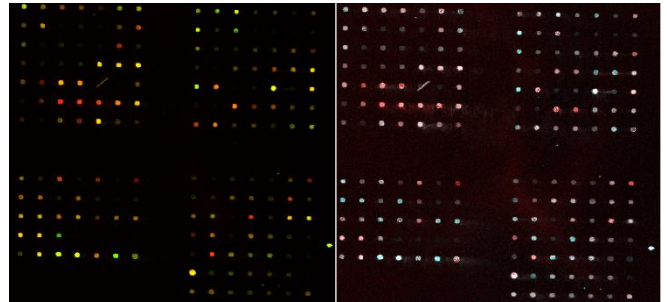


Figura 9. La imatge de l'esquerra correspon a la imatge en color generada del parell d'imatges Deff666CyX.raw i la de la dreta a la mateixa imatge però amb totes les millores - combinació canals R,G,G; valors dels píxels x2; gamma 0,8 -.

Pel que fa al mòdul escrit en Bash que realitza la compressió, validació i visualització d'imatges en color RGB, s'ha fet ús de quatre conjunts d'imatges [23] on era possible reconèixer a quines imatges se li havia aplicat la fluorescència Cy5 -vermella- i a quines la Cy3 -verda: apoa1, isrec, ibb i arizona. En la Taula 1 apareixen les característiques de cada conjunt.

TAULA 1  
CONJUNT D'IMATGES UTILITZAT

Característiques	apoa1	isrec	ibb	arizona
Any	2001	2001	2010	2011
Imatges	32	14	44	6
Mida	1044 x 1041	1000 x 1000	2019 x 6235	4400 x 13800
Disseny del spot	Quadrat	Quadrat	Quadrat	Hexagonal
Nombre de spots	$\sim 6 \cdot 10^5$	$\sim 2 \cdot 10^5$	$\sim 1,4 \cdot 10^4$	$\sim 2 \cdot 10^5$
Intensitat mitjana	39.51%	33.34%	6,09%	82.82%
Entropia mitjana	11.033	10.435	8,50	9.306

Per a cadascuna de les imatges de cadascun dels conjunts d'imatges s'han utilitzat diferents combinacions de nivells de transformades espacials i espectrals. La transformada espacial redueix la redundància espacial present dins de la imatge, amb la transformació realitzada cap a l'element vertical (files) i horitzontal (columnes) del domini espacial. D'altra banda, la transformada espectral transforma el domini espectral de la imatge per reduir la seva grandària. Els rangs utilitzats per al projecte són de 0 a 5 nivells espacials i de 0 a 4 nivells espectrals, deixant en aquest apartat les dues combinacions que millors resultats han aconseguit: 0 nivells espectrals juntament amb 5 nivells espacials, i 1 nivell espectral juntament amb 5 nivells espacials.

La Taula 2 conté la mitjana aritmètica de les dades obtingudes en la compressió per a cadascuna de les imatges corresponents a cada conjunt d'imatges amb 0 nivells espectral i 5 espacials. Aquestes dades són la mida

de la imatge, el nombre de bits per píxel (bpp) en la imatge - quocient de la divisió entre la mida de la imatge en bits i el producte de l'alçada, amplada i nombre de components de la imatge- i el Compress Ratio (CR) - quocient de la divisió entre la mida original de la imatge sense comprimir i una vegada comprimida -. Abans d'analitzar els resultats d'aquesta taula és necessari indicar que la mida de la imatge original està calculada com el sumatori de la mida dels dos components d'entrada - corresponents a la parella d'imatges en escala de grisos - i que el nombre de bpp de la imatge original només té en compte els de la parella d'imatges de microarray d'ADN, ja que la imatge nul·la per al canal blau no aporta cap informació en ser el valor de tots els seus píxels zero.

Es pot apreciar que el CR mitjà de tot el conjunt d'imatges és de 1,544280, la qual cosa equival a una compressió aproximada del 100 % -  $(1/1,544280) = 100\% - 0,647551 = 100\% - 64,7551\% = 35,2449\%$  i un bpp de **20,920130**. Aquesta és actualment la combinació que millor resultats ha donat.

TAULA 2  
RESULTATS TOTALS AMB 0 NIVELLS ESPECTRALS I 5 NIVELLS ESPACIALS

Nom conjunt	Mida Imatge original (Bytes)	Mida imatge comprimida (Bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
apoa1	4.353.084,00000	2.988.270,562500	32,000000	21,996758	1,457549
isrec	4.000.000,00000	2.828.115,428571	32,000000	22,624923	1,420503
ibb	50.405.886,818182	32.939.157,954546	32,000000	20,932915	1,533501
arizona	242.880.000,000000	137.575.764,333333	32,000000	18,125924	1,765565
Mitja conjunts	75,409.742,704546	44.082.827,069738	32,000000	20,920130	1,544280

En la Taula 3 es troben els resultats de la mitjana de tots els sets una vegada realitzada la compressió, igual que apareixia en la Taula 2, però en aquest cas aplicant 1 nivell de transformada espectral. Es pot apreciar que el CR mitjà de tot el conjunt d'imatges és de 1,044746, la qual cosa equival a una compressió aproximada del 100 % -  $(1/1,044746) = 100\% - 0,957170 = 100\% - 95,7170\% = 4,2830\%$  i un bpp de **30,922744**. Aquesta es la segona combinació amb millors resultats de compressió.

TAULA 3  
RESULTATS TOTALS AMB 1 NIVELL ESPECTRAL I 5 NIVELLS ESPACIALS.

Nom conjunt	Mida Imatge original (Bytes)	Mida imatge comprimida (Bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
apoa1	4.353.084,00000	4.361.831,062500	32,000000	32,107582	0,998615
isrec	4.000.000,00000	4.209.208,714286	32,000000	33,673670	0,954716
ibb	50.405.886,818182	48.810.941,090909	32,000000	31,019471	1,035098
arizona	242.880.000,000000	204.097.023,000000	32,000000	26,890253	1,190552
Mitja conjunts	75,409.742,704546	65.369.750,966924	32,000000	30,922744	1,044746

Analitzant els dos millors casos, es pot dir que valors CR tan baixos i bpp tan alts en el segon millor cas - 1 nivell espectral- indiquen un rendiment molt pobre en la compressió, fins al punt que en determinats conjunts d'imatges com isrec i apoa1, no només no es comprimeixen els bpp de la imatge sinó que el seu nombre augmenta de **32 bpp a 33,673670 bpp** i de **32 bpp a 32,107582 bpp**, respectivament. El principal problema d'aquest baix rendiment en la compressió es troba en l'ús de la

transformada espectral DWT 5/3 en tenir entre les tres imatges d'entrada del compressor una amb tots els seus píxels a zero, la qual contrasta amb la similitud existent entre la parella d'imatges en escala de grisos. S'ha comprovat que l'augment a 2, 3 i 4 nivells espectrals augmenta el nombre de bpp i mida de la imatge comprimida, per la raó explicada prèviament. No obstant això, en evitar l'ús de la transformada espectral especificant 0 nivells de transformada, s'incrementa notablement l'eficiència de la compressió.

Tenint en compte aquesta clara diferència en rendiment en fer o no ús de la transformada espectral DWT 5/3 i que en les proves realitzades en el document [5] amb aquesta i altres transformades espectrals com POT, DPCM o R-Haar -en els mateixos conjunts d'imatges que en aquest projecte, exceptuant el conjunt Yeast i amb l'addició del ibb- s'arribava a la conclusió que cap de les transformades espectrals esmentades era capaç de millorar significativament el rendiment de JPEG 2000; s'ha evitat l'ús de qualsevol d'aquestes. Un altre factor que ha millorat el CR i bitrate en la majoria dels conjunts d'imatges és l'ús de 5 nivells espacials, comprovant una millora creixent - encara que poc significativa - des de 0 nivells espacials a 5 i cap de 5 en endavant. En alguns conjunts d'imatges com isrec i ibb, l'ús de 0 nivells espacials produïa millors resultats que amb 5 nivells, però al no ser una diferència molt important i que amb la resta de conjunts no succeïa, s'ha acabat fixant el seu valor en 5 nivells espacials per tots el conjunts d'imatges. L'últim factor que ha millorat la compressió és l'ús del perfil ICC -explicat a l'apartat 2.Desenvolupament d'aquest article- ja que d'una altra forma, a més de no poder modificar el nivell gamma de la imatge amb Kakadu, s'hauria de realitzar el procés de combinació dels canals afegint en l'entrada del compressor l'ordre corresponent de les imatges i una tercera imatge que seria la copia de la imatge Cy3 o de la Cy5. Això implicaria un increment en el CR i bitrate, ja que Kakadu no és capaç de detectar que una de les imatges introduïdes per a un determinat canal està sent utilitzada al mateix temps per a un altre, tractant-la com una imatge més, en comptes de deixar aquest canal nul i renderitzar la imatge en color amb la combinació desitjada com si realitza el perfil ICC.

La matriu especificada A, comentada a l'apartat 2.Desenvolupament, s'ha definit a partir de diverses proves amb cadascuna de les imatges de tots els conjunts. S'ha dissenyat en base al fet que la imatge associada al canal vermell té un gran pes en el color resultant en generar la imatge en color RGB, així com el canal verd és el segon amb més pes, i el canal blau bàsicament aporta lluminositat a la imatge. És tal la rellevància del canal vermell a la imatge, que si aquest és una imatge amb tots els píxels a zero, la imatge en color RGB resultant tendeix al color negre absolut, encara que els altres dos canals tinguin associats imatges amb valors no nuls. Per aquesta raó s'ha optat per la combinació de (R, G, G), amb la finalitat de compensar la gran rellevància del canal vermell alhora que s'aporta més lluminositat a la imatge. Amb



aquesta configuració s'ha pogut distingir sense dificultat les diferents tonalitats aparegudes en generar les imatges en color RGB dels diversos conjunts d'imatges.

$$A = \begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 0.75 & 0 \\ 0 & 0.35 & 0 \end{pmatrix}$$

#### 4 CONCLUSIONS

En aquest apartat s'exposen les aportacions realitzades a aquest projecte com a TFG, els temes no tractats i les possibles extensions que podrien sorgir d'aquest treball. Les aportacions realitzades al TFG són tres:

- Una aplicació per a la generació, visualització i emmagatzematge d'imatges en color RGB de microarrays d'ADN a partir d'imatges en escala de grisos del mateix àmbit.
- Una aplicació que permet comprimir, validar i visualitzar imatges en color RGB de microarrays d'ADN
- Una revisió de l'estat de l'art en compressió de microarrays d'ADN dels últims anys, amb els mètodes més prometedors i els que millors resultats han obtingut en uns altres treballs.

La principal conclusió que es pot extreure d'aquest projecte és que amb l'aplicació per a la generació, visualització i emmagatzematge d'imatges en color RGB de microarrays d'ADN i l'aplicació per comprimir aquest mateix tipus d'imatges, l'usuari té al seu abast unes eines dotades de diferents utilitats per poder analitzar més fàcilment les imatges de microarrays d'ADN i guardar-les en un menor espai amb aquestes noves característiques.

En el que respecta la aportació de la aplicació per a la generació, visualització i emmagatzematge d'imatges, les aportacions que més destacaria són:

- La càrrega d'imatges en escala de grisos multiformat: RAW, PGM, PNG, TIF, JPG o JPEG. Evitant a l'usuari el problema d'haver de convertir les imatges de microarray d'ADN a altres formats per no ser compatibles amb l'aplicació.
- Llibertat a l'usuari perquè pugui combinar els canals que formaran la imatge en color RGB de la manera que li resulti més favorable a aquest, podent també compensar algun petit grau de daltonisme o dificultat per al reconeixement de tonalitats de colors.
- Comprovar cada imatge carregada/generada en qualsevol moment mitjançant l'ús de pestanyes, per la qual cosa es poden comprovar visualment les imatges en escala de grisos abans de realitzar la combinació, i decidir si es realitza o potser es prefereix carregar altres imatges.
- Modificar del nivell gamma de la imatge, fent aquesta més o menys fosca als ulls de l'usuari en temps real - sense necessitat de tornar a generar la imatge - i permetent-li ajustar aquest nivell a les seves necessitats.

- Millorar la distinció de les tonalitats de color de cada spot, una vegada generada la imatge en color, en multiplicar per dos el valor dels píxels que conformen els tres components de la imatge.
- Introdueix la possibilitat de guardar aquesta imatge en color en el sistema amb el format sense pèrdua TIFF (tif). És una forma de poder tenir guardats en el sistema diferents conjunts de combinacions de canals d'un mateix parell d'imatges de microarray d'ADN.

Per la aplicació que permet comprimir, validar i visualitzar imatges en color RGB de microarrays d'ADN. Les seves característiques més importants destinades a l'ús per part de l'usuari són:

- Genera una imatge en color amb una parella d'imatges en escala de grisos de microarray d'ADN amb els mateixos formats que l'aplicació de generació d'imatges en color, alhora que comprimeix la seva mida, amb només indicar la ruta de les imatges i les seves dimensions.
- Redueix la quantitat d'espai necessari per emmagatzemar les imatges de microarray d'ADN comprimint cada parella en una sola imatge en color. El que implica que en comptes de guardar N imatges per a un conjunt d'imatges, ara tan sols serà necessari guardar N/2 imatges, amb una mida menor al que ocupaven cada parella i podent recuperar aquestes imatges originals amb la realització del procés de descompressió d'aquestes imatges en color.
- Genera una imatge en color RGB amb una combinació de canals específica que permet identificar fàcilment les tonalitats de cada spot en la imatge -aclarir que aquesta combinació ha estat triada des d'un punt de vista subjectiu al desenvolupador -.

A continuació s'exposen els temes no tractats i les possibles extensions del treball realitzat:

- No s'ha tractat el tema de fer ús d'altres espais de color diferents al RGB, com podria ser l'espai de color YCbCr, el qual és una codificació no lineal de l'espai RGB. El paràmetre Y indica la luminància, els paràmetres Cb i Cr indiquen la tonalitat del color: Cb situa el color en una escala entre el blau i el groc, Cr indica la ubicació del color entre el vermell i el verd. És possible que l'espai de color YCbCr o uns altres que facin ús de més de tres components puguin o no millorar la compressió d'imatges en color de microarrays d'ADN. Aquest tema podria ser una bona extensió a aquest projecte.
- Es podria tractar, a partir d'aquest projecte, la cerca de mètodes que permetin millorar la compressió més enllà de les transformades espectrals i posant menor atenció en la millor manera de visualitzar les imatges en color a l'usuari. Aquests altres mètodes podrien basar-se en els treballs de Battiatto [5], [16] i Neves [5],[17], [18].

## AGRAÏMENTS

M'agradaria agrair al meu tutor de projecte, Joan Serra Sagristà, per tota l'ajuda proporcionada durant la realització del projecte; a Miguel Hernández Cabronero pel seu suport en la visualització i compressió de microarrays d'ADN; i a la meua família pel seu constant suport.

## BIBLIOGRAFIA

- [1] S. Lonardi i Y. Luo, "Gridding and Compression of Microarray Images." In Proceedings of the IEEE Computational Systems Bioinformatics Conference, pp. 122-130, 2004. Disponible: [http://www.cs.ucr.edu/~stelo/papers/CSB04\\_lonardi.pdf](http://www.cs.ucr.edu/~stelo/papers/CSB04_lonardi.pdf) . Últim accés: Febrer 2015.
- [2] Y. Zhang, R. Parthe, D. Adjeroh, "Lossless compression of DNA microarray images." In Computational Systems Bioinformatics Conference, 2005. Workshops and Poster Abstracts. IEEE, pp. 128-132, 2005. Disponible: [http://www.lifesciencesociety.org/CSB2005/PDF2/053\\_zhangy\\_microarray.pdf](http://www.lifesciencesociety.org/CSB2005/PDF2/053_zhangy_microarray.pdf) . Últim accés: Febrer 2015.
- [3] O. Gundogdu i A. Elmi, "Microarray Overview." In Genome Resource Facility (GRF) website, London School of Hygiene & Tropical Medicine. Disponible: <http://grf.lshtm.ac.uk/microarrayoverview.htm> . Últim accés: Febrer 2015.
- [4] N. Karimi, S. Samavi, S. Shirani, A. Banaei, E. Nasr-Esfahani, "Real-time lossless compression of microarray images by separate compaction of foreground and background." In Computer Standards & Interfaces, vol. 39, pp. 34-43, 2015. Disponible: <http://www.sciencedirect.com/science/article/pii/S0920548914001196> . Últim accés: Abril 2015.
- [5] M. Hernandez-Cabronero, M. W. Marcellin, J. Serra-Sagrista, "Compression of DNA Microarray Images." In Book "Microarray Image and Data Analysis: Theory and Practice.", CRC Press, pp. 193-225, 2014.
- [6] V. Moreno i X. Solé, "Uso de chips de ADN (microarrays) en medicina: fundamentos técnicos y procedimientos básicos para el análisis estadístico de resultados." En Medicina Clínica, vol. 122 (Supl. 1), pp. 73-79, 2004. Disponible: <http://www.sc.ehu.es/ccwubayes/docencia/mmcc/docs/divulgativos/UsodeChipsDeADN.pdf> . Últim accés: Febrer 2015
- [7] Genetic Science Learning Center, "DNA Microarray." Video interactiu. University of Utah, 2015. Disponible: <http://learn.genetics.utah.edu/content/labs/microarray/> . Últim accés: Febrer 2015.
- [8] J. González Maestre, "Análisis de datos de microarrays." Treball Final de Carrera en Enginyeria Informàtica. Universidad Politécnica de Valencia i Ibime (Informàtica Biomèdica), pp. 1-54, 2010. Disponible: <https://riunet.upv.es/bitstream/handle/10251/8578/Memoria%20PFC%20Jose%20Gonzalez.pdf?sequence=1> . Últim accés: Febrer 2015.
- [9] Institute of Biostatistics and Analyses, Masaryk University, "2.2.1.8 Analysis of arrayCGH." Apartat 2.2.1 "DNA microarrays" i subapartats. Disponible: <http://telemedicina.med.muni.cz/genomic-proteomic-analysis/index-en.php?pg=E-learning--2-data-analysis--2-2-analysis-of-genomic-data--2-2-1-DNA-microarrays-1--2-2-1-8-analysis-of-arraycgh> . Últim accés: Març 2015
- [10] R. M. Simon, E. L. Korn, Lisa M. McShane, M. D. Radmacher, G. W. Wright, Y. Zhao, "Design and Analysis of DNA Microarray Investigations." In Springer Science & Business Media, Capítol 4 (pp. 29-38), 2003. Disponible: <http://linus.nci.nih.gov/techreport/DesignandAnalysisofDNA-MicroarrayInvestigations.pdf> . Últim accés: Març 2015.
- [11] ArgenBio. Consejo Argentino para la Información y el Desarrollo de la Biotecnología. Disponible: [http://www.argenbio.org/adc/uploads/imagenes\\_doc/omics/transcriptomica.bmp](http://www.argenbio.org/adc/uploads/imagenes_doc/omics/transcriptomica.bmp) . Últim accés: Maig 2015.
- [12] National Electrical Manufacturers Association, "Digital Image and Communication in Medicine standard, DICOM." Disponible: <http://medical.nema.org/> . Últim accés: Març 2015.
- [13] M. Rabbani i R. Joshi, "An overview of the JPEG 2000 still image compression standard." In Signal processing: Image communication, vol. 17, nº 1, pp. 3-48, 2002.
- [14] R. Jornsten, W. Wang, B. Yu, K. Ramchandran, "Microarray image compression: SLOCO and the effect of Information loss." In Genomic Signal Processing, vol. 83, issue 4, pp. 859-869, 2003. Disponible: <http://www.sciencedirect.com/science/article/pii/S016516840204784> . Últim accés: Març 2015.
- [15] J. Hua, Z. Liu, Z. Xiong, Q. Wu i K. R. Castleman, "Microarray BASICA: background adjustment, segmentation, image compression and analysis of microarray images." In EURASIP Journal on Applied Signal Processing, pp. 92-107, 2004. Disponible: <http://dl.acm.org/citation.cfm?id=1289347> . Últim accés: Abril 2015.
- [16] S. Battiato and F. Rundo, "A bio-inspired CNN with re-indexing engine for lossless DNA microarray compression and segmentation," in Proceedings of the 16th International Conference on Image Processing, vol. 1-6, pp. 1717-1720, 2009.
- [17] A. J. R. Neves and A. J. Pinho, "Lossless compression of microarray images," in Proceedings of the International Conference on Image Processing, ICIP, pp. 2505-2508, 2006.
- [18] A. J. R. Neves and A. J. Pinho, "Lossless compression of microarray images using image-dependent finite-context models," IEEE Transactions on Medical Imaging, vol. 28, pp. 194-201, February 2009.
- [19] J. H. Canós, P. Letelier, M. C. Penadés, Grupo ISSI (Ingeniería del Software y Sistema de Información), "Metodologías Ágiles en el Desarrollo de Software." En VIII Jornadas de Ingeniería del Software y Base de Datos, JISBD 2003. Disponible: <http://issi.dsic.upv.es/archives/f-1069167248521/actas.pdf> . Últim accés: Febrer 2015.
- [20] O. Tinoco Gómez, P. Pablo Rosales López, J. Salas Bacalla, "Criterios de selección de metodologías de desarrollo de software." Revista de la Facultad de Ingeniería Industrial, vol. 13, nº 2, pp. 70-74, 2010. Disponible: <http://revistas.concytec.gov.pe/pdf/id/v13n2/a09v13n2.pdf> . Últim accés: Febrer 2015.
- [21] Departamento de Computación, "Preparación de Plan de Proyecto." Material asignatura "Ingeniería de Software II." En Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 2005. Disponible: [http://www-2.dc.uba.ar/materias/isoft2/2005\\_02/clases/PlanProyecto%28incWBS%2920050905.pdf](http://www-2.dc.uba.ar/materias/isoft2/2005_02/clases/PlanProyecto%28incWBS%2920050905.pdf) . Últim accés: Febrer 2015.
- [22] Kakadu Software version 7.4. Disponible: <http://kakadusoftware.com/> . Últim accés: Abril 2015.
- [23] M. Hernández-Cabronero, "Additional materials from Miguel Hernández-Cabronero." DNA microarray sets, 2015. Disponible: <http://deic.uab.es/~mhernandez/materials> . Últim accés: Abril 2015.

## APÈNDIX

### A1. RESULTATS DEL CONJUNT D'IMATGES APOA1 1044 x 1041 AMB 1 NIVELL ESPECTRAL I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (bytes)	Mida Imatge Comprimida (bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
1230c1R-RGB.jpg	4.353.084	4.601.617	32,000000	33,872654	0,945990
1230c2R-RGB.jpg	4.353.084	4.599.215	32,000000	33,854973	0,946484
1230c3R-RGB.jpg	4.353.084	4.323.925	32,000000	31,828554	1,006744
1230c4R-RGB.jpg	4.353.084	4.385.635	32,000000	32,282804	0,992578
1230c5R-RGB.jpg	4.353.084	4.240.632	32,000000	31,215432	1,026518
1230c6R-RGB.jpg	4.353.084	4.408.320	32,000000	32,449789	0,987470
1230c7R-RGB.jpg	4.353.084	4.296.980	32,000000	31,630211	1,013057
1230c8R-RGB.jpg	4.353.084	4.422.827	32,000000	32,556575	0,984231
1230ko1R-RGB.jpg	4.353.084	4.207.471	32,000000	30,971332	1,034608
1230ko2R-RGB.jpg	4.353.084	4.238.230	32,000000	31,197750	1,027100
1230ko3R-RGB.jpg	4.353.084	4.388.737	32,000000	32,305637	0,991876
1230ko4R-RGB.jpg	4.353.084	4.271.895	32,000000	31,445560	1,019005
1230ko5R-RGB.jpg	4.353.084	4.389.103	32,000000	32,308332	0,991794
1230ko6R-RGB.jpg	4.353.084	4.301.004	32,000000	31,659832	1,012109
1230ko7R-RGB.jpg	4.353.084	4.366.993	32,000000	32,145579	0,996815
1230ko8R-RGB.jpg	4.353.084	4.346.713	32,000000	31,996297	1,001466
Mitja set	4.353.084,000000	4.361.831,062500	32,000000	32,107582	0,998615

### A2. RESULTATS DEL CONJUNT D'IMATGES ISREC 1000 x 1000 AMB 1 NIVELL ESPECTRAL I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (bytes)	Mida Imatge Comprimida (bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
Def661Cy5.tif.pgm-RGB.jpg	4.000.000	4.202.612	32,000000	33,620896	0,951789
Def662Cy5.tif.pgm-RGB.jpg	4.000.000	4.030.999	32,000000	32,247992	0,992310
Def663Cy5.tif.pgm-RGB.jpg	4.000.000	4.225.761	32,000000	33,806088	0,946575
Def664Cy5.tif.pgm-RGB.jpg	4.000.000	4.261.162	32,000000	34,089296	0,938711
Def665Cy5.tif.pgm-RGB.jpg	4.000.000	4.834.320	32,000000	38,674560	0,827417
Def666Cy5.tif.pgm-RGB.jpg	4.000.000	4.115.688	32,000000	32,925504	0,971891
Def667Cy5.tif.pgm-RGB.jpg	4.000.000	3.793.919	32,000000	30,351352	1,054319
32,000000	4.000.000,000000	4.209.208,714286		33,673670	0,954716

### A3. RESULTATS DEL CONJUNT D'IMATGES IBB 2019 x 6235 AMB 1 NIVELL ESPECTRAL I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (bytes)	Mida Imatge Comprimida (bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
134044018_Cy5_635-RGB.jpg	50.405.888	50.805.590	32,000000	32,287076	0,992133
134044019_Cy5_635-RGB.jpg	50.405.886	48.594.435	32,000000	30,881881	1,037277
134044022_Cy5_635-RGB.jpg	50.405.886	48.451.388	32,000000	30,790974	1,040339
134044023_Cy5_635-RGB.jpg	50.405.886	52.435.677	32,000000	33,323000	0,961290
Cy5_14106171_635-RGB.jpg	50.405.890	51.505.637	32,000000	32,731957	0,978648
Cy5_14347649_635-RGB.jpg	50.405.890	50.254.434	32,000000	31,936815	1,003014

Cy5_ref24h_635-RGB.jpg	50.405.886	49.024.924	32,000000	31,155458	1,028169
Cy5_ref28h_635-RGB.jpg	50.405.886	49.161.667	32,000000	31,242358	1,025309
Cy5_wt8h_635-RGB.jpg	50.405.886	48.760.634	32,000000	30,987501	1,033741
Cy5_wt24h_635-RGB.jpg	50.405.886	48.334.155	32,000000	30,716473	1,042863
ref232h_Cy5_635-RGB.jpg	50.405.886	51.135.703	32,000000	32,496863	0,985728
W303_Cy5_635-RGB.jpg	50.405.886	50.923.145	32,000000	32,361782	0,989842
wt32h_Cy5_635-RGB.jpg	50.405.886	52.153.898	32,000000	33,143929	0,966484
Y_all_16px2_36_14164789_Cy5_635-RGB.jpg	50.405.886	47.519.012	32,000000	30,198447	1,060752
Y_all_16px2_36_14164789_Cy5_2_635-RGB.jpg	50.405.886	46.629.426	32,000000	29,633113	1,080989
Y_all_16px2_36_14165006_Cy5_635-RGB.jpg	50.405.888	45.438.568	32,000000	28,876320	1,109319
Y_all_16px2_36_14165006_Cy5_2_635-RGB.jpg	50.405.888	43.532.029	32,000000	27,664710	1,157903
Y_all_16px2_36_14165010_Cy5_635-RGB.jpg	50.405.886	47.852.384	32,000000	30,410306	1,053362
Y_all_16px2_36_14165010_Cy5_2_635-RGB.jpg	50.405.888	47.133.083	32,000000	29,953188	1,069438
Y_all_16px2_36_14165074_Cy5_635-RGB.jpg	50.405.886	46.531.989	32,000000	29,571192	1,083252
Y_all_16px2_36_14165074_Cy5_2_635-RGB.jpg	50.405.888	45.830.155	32,000000	29,125175	1,099841
YWK186_Cy5_635-RGB.jpg	50.405.886	51.832.771	32,000000	32,939852	0,972471
Mitja set	50.405.886,818182	48.810.941,090909	32,000000	31,019471	1,035098

#### A4. RESULTATS DEL CONJUNT D'IMATGES ARIZONA 4400 X 13800 AMB 1 NIVELL ESPECTRAL I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (bytes)	Mida Imatge Comprimida (bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
slide_1-red.tif.pgm-RGB.tif	242.880.000	201.311.232	32,000000	26,523219	1,206490
slide_2-red.tif.pgm-RGB.tif	242.880.000	200.755.314	32,000000	26,449975	1,209831
slide_3-red.tif.pgm-RGB.tif	242.880.000	210.224.523	32,000000	27,697566	1,155336
Mitja set	242.880.000,000000	204.097.023,000000	32,000000	26,890253	1,190552

#### A5. RESULTATS DEL CONJUNT D'IMATGES APOA1 1044 X 1041 AMB 0 NIVELLS ESPECTRALS I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (Bytes)	Mida Imatge Comprimida (Bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
1230c1R-RGB.jpg	4.353.084	3.154.344	32,000000	23,219230	1,380028
1230c2R-RGB.jpg	4.353.084	3.138.433	32,000000	23,102109	1,387025
1230c3R-RGB.jpg	4.353.084	2.946.017	32,000000	21,685728	1,477617
1230c4R-RGB.jpg	4.353.084	3.000.304	32,000000	22,085336	1,450881
1230c5R-RGB.jpg	4.353.084	2.896.980	32,000000	21,324765	1,502628
1230c6R-RGB.jpg	4.353.084	3.022.338	32,000000	22,247529	1,440304
1230c7R-RGB.jpg	4.353.084	2.915.125	32,000000	21,458331	1,493275
1230c8R-RGB.jpg	4.353.084	3.045.015	32,000000	22,414456	1,429577
1230ko1R-RGB.jpg	4.353.084	2.925.220	32,000000	21,532641	1,488122
1230ko2R-RGB.jpg	4.353.084	2.920.055	32,000000	21,494621	1,490754

1230ko3R-RGB.jpg	4.353.084	3.003.590	32,000000	22,109525	1,449294
1230ko4R-RGB.jpg	4.353.084	2.953.573	32,000000	21,741348	1,473837
1230ko5R-RGB.jpg	4.353.084	2.993.738	32,000000	22,037004	1,454063
1230ko6R-RGB.jpg	4.353.084	2.938.131	32,000000	21,627679	1,481583
1230ko7R-RGB.jpg	4.353.084	2.978.815	32,000000	21,927155	1,461348
1230ko8R-RGB.jpg	4.353.084	2.980.651	32,000000	21,940670	1,460447
<b>Mitja set</b>	<b>4.353.084,000000</b>	<b>2.988.270,562500</b>	<b>32,000000</b>	<b>21,996758</b>	<b>1,457549</b>

## A6. RESULTATS DEL CONJUNT D'IMATGES ISREC 1000 x 1000 AMB 0 NIVELLS ESPECTRALS I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (Bytes)	Mida Imatge Comprimida (Bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
Def661Cy5.tif.pgm-RGB.jpg	4.000.000	2.690.643	32,000000	21,525144	1,486633
Def662Cy5.tif.pgm-RGB.jpg	4.000.000	2.767.994	32,000000	22,143952	1,445090
Def663Cy5.tif.pgm-RGB.jpg	4.000.000	2.774.914	32,000000	22,199312	1,441486
Def664Cy5.tif.pgm-RGB.jpg	4.000.000	2.907.040	32,000000	23,256320	1,375970
Def665Cy5.tif.pgm-RGB.jpg	4.000.000	3.242.482	32,000000	25,939856	1,233623
Def666Cy5.tif.pgm-RGB.jpg	4.000.000	2.821.104	32,000000	22,568832	1,417885
Def667Cy5.tif.pgm-RGB.jpg	4.000.000	2.592.631	32,000000	20,741048	1,542834
<b>Mitja set</b>	<b>4.000.000,000000</b>	<b>2.828.115,428571</b>	<b>32,000000</b>	<b>22,624923</b>	<b>1,420503</b>

## A7. RESULTATS DEL CONJUNT D'IMATGES IBB 2019 x 6235 AMB 0 NIVELLS ESPECTRALS I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (Bytes)	Mida Imatge Comprimida (Bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
134044018_Cy5_635-RGB.jpg	50.405.888	34.707.690	32,000000	22,056821	1,452297
134044019_Cy5_635-RGB.jpg	50.405.886	33.240.060	32,000000	21,124139	1,516420
134044022_Cy5_635-RGB.jpg	50.405.886	33.153.439	32,000000	21,069091	1,520382
134044023_Cy5_635-RGB.jpg	50.405.886	34.866.763	32,000000	22,157912	1,445672
Cy5_14106171_635-RGB.jpg	50.405.890	34.598.348	32,000000	21,987334	1,456887
Cy5_14347649_635-RGB.jpg	50.405.890	33.765.927	32,000000	21,458328	1,492803
Cy5_ref24h_635-RGB.jpg	50.405.886	32.944.171	32,000000	20,936100	1,530040
Cy5_ref28h_635-RGB.jpg	50.405.886	32.934.798	32,000000	20,930144	1,530475
Cy5_wt8h_635-RGB.jpg	50.405.886	32.817.224	32,000000	20,855425	1,535958
Cy5_wt24h_635-RGB.jpg	50.405.886	31.996.650	32,000000	20,333949	1,575349
ref232h_Cy5_635-RGB.jpg	50.405.886	33.861.314	32,000000	21,518947	1,488598
W303_Cy5_635-RGB.jpg	50.405.886	34.374.395	32,000000	21,845011	1,466379
wt32h_Cy5_635-RGB.jpg	50.405.886	35.173.246	32,000000	22,352683	1,433075
Y_all_16px2_36_14164789_Cy5_635-RGB.jpg	50.405.886	32.006.330	32,000000	20,340100	1,574872
Y_all_16px2_36_14164789_Cy5_2_635-RGB.jpg	50.405.886	31.496.015	32,000000	20,015794	1,600389
Y_all_16px2_36_14165006_Cy5_635-RGB.jpg	50.405.888	31.033.042	32,000000	19,721573	1,624265
Y_all_16px2_36_14165006_Cy5_2_635-RGB.jpg	50.405.888	29.761.861	32,000000	18,913735	1,693640
Y_all_16px2_36_14165010_Cy5_635-RGB.jpg	50.405.886	32.302.138	32,000000	20,528087	1,560450

Y_all_16px2_36_14165010_Cy 5_2_635-RGB.jpg	50.405.888	32.041.091	32,000000	20,362191	1,573164
Y_all_16px2_36_14165074_Cy 5_635-RGB.jpg	50.405.886	31.228.560	32,000000	19,845826	1,614096
Y_all_16px2_36_14165074_Cy 5_2_635-RGB.jpg	50.405.888	30.979.612	32,000000	19,687618	1,627066
YWK186_Cy5_635-RGB.jpg	50.405.886	35.378.801	32,000000	22,483314	1,424748
Mitja set	50.405.886,818182	32.939.157,954546	32,000000	20,932915	1,533501

#### A8. RESULTATS DEL CONJUNT D'IMATGES ARIZONA 4400 X 13800 AMB 0 NIVELLS ESPECTRALS I 5 NIVELLS ESPACIALS

Nom imatge	Mida Imatge original (Bytes)	Mida Imatge Comprimida (Bytes)	Bit píxel imatge original (bpp)	Bit píxel imatge comprimida (bpp)	CR
slide_1-red.tif.pgm-RGB.tif	242.880.000	136.828.200	32,000000	18,027431	1,775073
slide_2-red.tif.pgm-RGB.tif	242.880.000	136.602.447	32,000000	17,997687	1,778006
slide_3-red.tif.pgm-RGB.tif	242.880.000	139.296.646	32,000000	18,352654	1,743617
Mitja set	242.880.000,000000	137.575.764,333333	32,000000	18,125924	1,765565