

Instituto Tecnológico de Costa Rica  
Vicerrectoría de Investigación y Extensión  
Dirección de Proyectos

Informe Final de Proyecto de Investigación  
Documento II  
Periodo 2010–2011

**Análisis por computador de imágenes de geles de electroforesis:  
métodos avanzados de manejo de meta-información y  
procesamiento digital de imágenes**

5402 1360 2601

Adscrito a:  
Escuela de Ingeniería Electrónica

Investigador principal:  
Dr. José Pablo Alvarado, EIE

Investigadores:  
M.Sc. Alicia Salazar, CIC    Dr. José Enrique Araya, CIC    M.Sc. Johnny Peraza, CIB  
Ing. Fabiana Rojas, CIIBI    Dr. Olman Murillo, CIIBI

Estudiantes tesisistas:  
Pedro Alpízar Salas    Bryant Álvarez Canales    Pablo Barrantes Chaves  
Randall Esquivel Alvarado    Edison Fernández Alvarado    David Soto Vásquez

4 de julio, 2012



# Índice general

1	Datos Generales	1
2	Grado de avance	3
3	Limitaciones	11
4	Observaciones generales y recomendaciones	13
	Bibliografía	15



# Capítulo 1

## Datos Generales

**Nombre del proyecto:** Análisis por computador de imágenes de geles de electroforesis: métodos avanzados de manejo de meta-información y procesamiento digital de imágenes

**Código:** 5402 1360 2601

**Departamento académico responsable:** Escuela de Ingeniería Electrónica

**Investigador responsable:** Dr. Pablo Alvarado Moya

**Otros investigadores:**

Ing. Alicia Salazar	CIC
Dr. José Enrique Araya	CIC
Dr. Olman Murillo	CIIBI
Ing. Fabiana Rojas	CIIBI
M.Sc. Johnny Peraza	CIB

**Periodo cubierto:** 2010–2011



# Capítulo 2

## Grado de avance

El objetivo general del presente proyecto ha sido incorporar a la herramienta desarrollada en los proyectos antecesores [4, 5] opciones avanzadas para el manejo de meta-información y para el análisis de imágenes de geles de electroforesis utilizadas en la caracterización molecular de organismos, de modo que se mejore la usabilidad del sistema en laboratorios de biología molecular.

Los objetivos planteados en la propuesta original ante la Vicerrectoría de Investigación y Extensión fueron:

1. Mejorar la seguridad en el manejo de los datos del sistema por medio del concepto de “administración de roles”.
2. Extender la arquitectura del sistema para manejar bases de datos distribuidas.
3. Agregar a la interfaz de usuario modos adicionales para la manipulación de las imágenes, la interacción con el sistema y la presentación de la información almacenada y generada.
4. Diseñar e implementar un sistema integrado de captura de imágenes de geles de electroforesis, que sea seguro para los usuarios finales.
5. Identificar automáticamente los carriles de control con base en datos almacenados.
6. Crear un módulo integrado de rectificación y normalización de imágenes de geles.
7. Integrar información contextual en los algoritmos de análisis de imágenes de geles.
8. Integrar la meta-información, imágenes y métodos de análisis de diversidad genética bacteriana en muestras ambientales utilizando la DGGE.

El ajuste en la asignación de presupuesto requirió que dichos objetivos fueran modificados de la siguiente forma:

Objetivo 1: Se reduce a un 70%, concluyendo con la realización de un prototipo preliminar, en vez de perseguir la implementación completa.

Objetivo 2: Se reduce a un 70%, aspirando a un diseño preliminar de la arquitectura distribuida.

Objetivo 3: Se mantiene al 100%, por ser el núcleo central en el desarrollo de una plataforma utilizable de forma eficiente en el laboratorio.

Objetivo 4: Se elimina. Este objetivo requería los fondos de equipo, que fue necesario reemplazar por presupuesto para asistentes. Al ser eliminado, en el Documento I no se menciona.

Objetivo 5: Se mantiene al 100%.

Objetivo 6: Se reduce a un 70%.

Objetivo 7: Se reduce a un 50%.

Objetivo 8: Se reduce a un 90%, pues se planearon secciones relacionadas con el desarrollo de la plataforma de captura que no serán posibles al ser eliminado el objetivo 4.

De acuerdo a estos últimos, la tabla 2.1 ilustra el alcance del proyecto. Los porcentajes se refieren a la totalidad del proyecto. Promediando los avances de cada objetivo, se estima que el alcance total del proyecto es de un 97%. La razón principal del incumplimiento total se atribuye al lento avance en la parte de informática, para la cual se contrataron en el último año asistentes de electrónica que han debido aprender gran cantidad de conceptos ajenos a sus áreas en una curva de aprendizaje más abrupta de lo que inicialmente se estimó. Recuérdese que debido a la enorme dificultad en semestres anteriores de conseguir estudiantes asistentes de la carrera de computación interesandos en laborar en el proyecto fue que se tomó esta decisión de intentar cubrir las tareas con estudiantes de electrónica. Este problema empeoró con la finalización de los periodos de asistencia de los estudiantes, que fue de un año, requiriendo reentrenar nuevos estudiantes. Dos asistentes de Ingeniería en Computadores fueron contratados (Javier Montoya y Edward Jiménez), pero por sus elevadas cargas académicas, el tiempo dedicado al proyecto fue limitado. Esta parece ser la mejor opción a futuro, pues estos estudiantes tienen mejores bases en el área de computación que los estudiates de la carrera de electrónica, sin embargo, para la última parte del proyecto fue imposible conseguir más asistentes de esta carrera.

En el contexto del proyecto se realizaron las tesis de licenciatura en ingeniería electrónica de

1. Bryant Álvarez [6] en el área de mejoramiento de calidad de imagen en la captura (aporta a los objetivos 3, 7 y 4).
2. Pablo Barrantes [7] en el área de detección de carriles (aporta a los objetivos 3, 5, 6 y 7)
3. Pedro Alpízar [3] quien hizo una propuesta muy novedosa de utilización de técnicas de difusión de imágenes para atacar el problema de la distorsión del efecto sonrisa, que complementa los trabajos realizados por Antonio Aguilar [1] en el proyecto anterior.
4. David Soto [12] en el área de detección de bandas por medios de optimización (aporta a los objetivos 5, 6 y 7)
5. Edison Fernández [10] complementa el trabajo de David Soto proponiendo un método para detectar el ancho de las bandas.



6. Randall Esquivel [9], que continuó con el tema concreto de detección de bandas iniciado por David Soto [12] y continuado por Edison Fernández [10]. Esta tesis ganó el primer lugar del Premio Asoelectrónica 2011 a los mejores Proyectos de Graduación de Ingeniería Electrónica.
7. José Ángel Umaña inició una tesis enfocada completamente en el 5 pero por razones personales no la concluyó.

Como asistentes han participado en el proyecto además de ingeniería en electrónica Mauricio Caamaño, Bayron Monge, Eduardo Corrales, Diego Sánchez y Andrés González; los estudiantes de ingeniería en computadores Javier Montoya y Edward Jiménez, y las estudiantes de ingeniería en biotecnología Shirleny Sandoval, Rossi Guillén y Milenna González.

En cuanto a divulgación, los resultados se presentaron en la conferencia TSD2011 (organizada por el programa doctoral DOCINADE) a finales de enero de 2011 con los artículos [2, 11, 8] y en el concurso Premio Asoelectrónica 2011 a los mejores Proyectos de Graduación de Ingeniería Electrónica con varias de las tesis.

Tabla 2.1: Avance de objetivos

Objetivo específico	Productos	Actividades	Fechas	Avance	Comentarios
1. Mejorar la seguridad en el manejo de los datos del sistema por medio del concepto de "administración de roles"	Módulo de administración de roles de usuarios	1.1. Análisis preliminar de arquitectura del sistema completo y el sistema de administración de roles 1.2. Diseño del módulo de administración de roles del sistema 1.3. Implementación del sistema 1.4. Verificación 1.5. Rediseño	Feb.10  Mar.10–Abr.10 May.10–Feb.11 Ene.11–Nov.11 Mar.11–Oct.11	100%	Se completó el análisis y arquitectura del sistema. En el módulo de administración de roles se completó la interfaz de creación de usuarios y compañías. Se reestructuraron algunos detalles por problemas a nivel de base de datos. Se tiene ya la validación del ingreso al sistema, por usuario y clave, que redirige a las páginas correctas en caso de éxito o no en el ingreso. Se reprogramaron algunas funciones del sistema como los mantenimientos de analista, el equipo, marcas, modelos. Se solucionaron problemas de links perdidos dentro de la aplicación.
2. Extender la arquitectura del sistema para manejar bases de datos distribuidas.	Sistema distribuido de manejo de bases de datos.	2.1. Análisis preliminar de arquitectura del sistema con respecto a la distribución del SABD 2.2. Diseño del sistema distribuido de BD 2.3. Implementación del sistema 2.4. Verificación y validación 2.5. Rediseño	Feb.10  Mar.10–May.10 Jun.10–Dic.10 Ene.11–Nov.11 Mar.11–Oct.11	100%	Por reducción presupuestaria el alcance original se redujo a un 70%, aspirando a un diseño preliminar de la arquitectura distribuida, el cual se concluyó. Se corrigieron las limitaciones en la plataforma con un ajuste del diseño del sistema distribuido propuesto inicialmente.

*Continúa en la siguiente página...*

Objetivo específico	Productos	Actividades	Fechas	Avance	Comentarios
<p>3. Agregar a la interfaz de usuario modos adicionales para la manipulación de las imágenes, la interacción con el sistema y la presentación de la información almacenada y generada.</p>	<p>Modos de uso en la herramienta</p>	<p>3.1.Levantamiento de lista de requisitos de módulos de uso 3.2.Sesiones de brainstorming con usuarios finales 3.3.Diseño del módulos de uso 3.4.Implementación de módulos 3.5.Verificación 3.6.Rediseño</p>	<p>Jul.10–Set.11 Ene.11–Nov.11 Mar.11–Oct.11</p>	<p>100% (módulos de seguridad y protocolos) 100% (minería de datos)</p>	<p>Se integró el uso de DBus al sistema para facilitar la comunicación entre PHP, Java y C++ utilizados por un lado en la plataforma web, y por otro en el sistema de captura y procesamiento de imágenes. La complejidad ha sido muy elevada, lo que ha reducido el avance esperado. El asistente Mauricio Caamaño se encargó de la programación de esta parte. Para el levantamiento de requisitos y diseño se tiene un 100%, para módulos de seguridad, protocolos y para minería de datos un 100%. Se identificaron y analizaron las funciones de minería de datos que se incorporaron al proyecto. Se modificó la BD para poder dar soporte al manejo de las imágenes y la información de carriles. Se desarrolló el mecanismo requerido para invocar las bibliotecas que realizan el procesamiento de análisis de datos. Se incorporaron operaciones de clustering y de clasificación. Los atrasos debido a la curva de aprendizaje para los asistentes se recuperó parcialmente con tiempo de los investigadores del área de computación, pero sus horas semanales fueron insuficientes para abarcar la cantidad de trabajo requerida.</p>
<p>4. Diseñar e implementar un sistema integrado de captura de imágenes de geles de electroforesis, que sea seguro para los usuarios finales.</p>				<p>—</p>	<p>Eliminado. (Objetivo eliminado desde la aceptación del presupuesto asignado)</p>
<p>5. Identificar automáticamente los carriles de control con base en datos almacenados.</p>	<p>Algoritmos de generalización de carriles y de búsqueda</p>	<p>5.1.Revisión bibliográfica sobre sistemas similares en funcionalidad. 5.2.Diseño de sistema de generalización y búsqueda de carriles de control. 5.3.Implementación del sistema 5.4.Verificación y validación 5.5.Rediseño</p>	<p>Jul.10–Ago.10 Ago.10–Dic.10 Ene.11–Jun.11 Abr.11–Nov.11 Jul.11–Set.11</p>	<p>85%</p>	<p>En la tesis de José Ángel Umaña (inconchusa) se exploró el uso de DTW (Dynamic Time Warping) en la identificación de carriles de control estándares. Surgió el problema de uso parcial de carriles de control comerciales, pero el algoritmo elegido logra parcialmente lidiar con ese problema, aunque otras alternativas deberán buscarse. Queda para futuros proyectos evaluar la extensión de este método a carriles no estándares, es decir, carriles que los biólogos utilicen como controles particulares de sus aplicaciones (una bacteria específica, ADN de un individuo particular, etc.)</p>

*Continúa en la siguiente página...*

Objetivo específico	Productos	Actividades	Fechas	Avance	Comentarios
6. Crear un módulo integrado de rectificación y normalización de imágenes de geles.	Algoritmo integrado de rectificación y normalización de imágenes	6.1.Revisión bibliográfica de sistemas integrados de rectificación en aplicaciones relevantes 6.2.Diseño del módulo integrado de rectificación 6.3.Implementación del módulo integrado de rectificación 6.4.Verificación del módulo 6.5.Rediseño	Ene.10-Feb.10  Mar.10-May.10  Jun.10-Set.10  Oct.10-Jun.11 Ene.11-Mar.11  Ene.11-Feb.11	100%	La combinación de las tesis de Pablo Barrantes (detección de carriles), y Pedro Alpizar (detección de efecto sonrisa) logra integrar la detección de carriles con la rectificación y corrección del efecto sonrisa.
7. Integrar información contextual en los algoritmos de análisis de imágenes de geles.	Algoritmos de mejoramiento de imágenes de geles. Aspectos contextuales en algoritmos de rectificación y normalización	7.1.Revisión bibliográfica de modelos matemáticos para los procesos de generación de carriles y bandas. 7.2.Diseño de al menos un algoritmo de mejoramiento de imagen 7.3.Adaptación de los algoritmos de rectificación y normalización a restricciones impuestas por la naturaleza de generación de los geles.	Mar.11-Jul.11  Mar.11-Nov.11	100%	Las tesis de David Soto y Edison Fernández fueron complementadas con la tesis de Randall Esquivel, que incorporan información sobre la forma de bandas para su mejor detección. Resultados de la tesis de Pedro Alpizar se extienden directamente a la detección de carriles (no solo líneas de bandas), y aumentar así la confiabilidad de los métodos explorados por Pablo Barrantes. La tesis de Bryant Álvarez también incorpora características propias de las imágenes de geles para mejorar la captura.

*Continúa en la siguiente página...*

Objetivo específico	Productos	Actividades	Fechas	Avance	Comentarios
<p>8. Integrar la meta-información, imágenes y métodos de análisis de diversidad genética bacteriana en muestras ambientales utilizando la DGGE.</p>	<p>Producción de geles DGGE. Módulos de captura, almacenamiento y análisis de datos de geles DGGE en plataforma.</p>	<p>8.1.Revisión bibliográfica. 8.2.Compra de equipo, materiales y reactivos de laboratorio. 8.3.Establecimiento del equipo y optimización del protocolo para la DGGE. 8.4.Colecta de muestras y extracción de ADN de comunidades microbianas. 8.5.Amplificación por PCR de genes ARNr. 8.6.Análisis por DGGE de los productos amplificados obtenidos de muestras ambientales. 8.7.Captura de imágenes de geles DGGE. 8.8.Adaptaciones del sistema de bases de datos a geles DGGE 8.9.Desarrollo de algoritmos de análisis de geles DGGE.</p>	<p>Feb.10-Mar.10 Mar.10-Jun.10  Mar.10-Dic.10  Jun.10-Dic.10  Jun10-Jun.11 Ene.11-Set.11  Mar.11-Set.11  Jul.10-Dic.10  Ago.11-Oct.11</p>	<p>100%</p>	<p>Concluido. Geles en DGGE fueron generados y el protocolo ya ha sido establecido. El sistema de base de datos es capaz de almacenar la información relacionada con este tipo de geles.</p>

Debió invertirse gran cantidad de tiempo re-escribiendo código para el manejo de las cámaras industriales utilizadas, pues la evolución natural del software libre hizo que el código escrito previamente para el proyecto quedara invalidado por la obsolescencia de la biblioteca subyacente (`libdc1394`). En esto se debió invirtió un asistente (Diego Sánchez) y el tiempo del investigador principal durante el segundo semestre de 2011, pues dicho código es altamente complejo. Los resultados han sido integrados a la biblioteca LTI-Lib-2, en la que se basan todos los trabajos presentados sobre procesamiento de imágenes.

Una razón no mencionada para el corto alcance del objetivo 1 es que el computador portátil del investigadora Alicia Salazar fue robado con gran parte del código corregido para el proyecto. Dicho código está siendo reconstruido en este momento.

# Capítulo 3

## Limitaciones

Más allá de las dificultades propias del tema de investigación, que han presentado retos en el diseño de algoritmos para el análisis de la imágenes inesperados, y algunos problemas personales de los estudiantes de licenciatura que imposibilitaron la culminación sus tesis (tesis de José Ángel Umaña), el problema fundamental para este proyecto ha sido encontrar estudiantes que colaboren con la programación avanzada del sistema. La cantidad de horas requerida para completar un prototipo integrado es muy elevada, y las horas disponibles para los investigadores principales hace imposible que se realicen todas las labores propias de investigación y se agreguen las horas requeridas de desarrollo e implementación.

Se contrató en la segunda etapa del proyecto a dos estudiantes de ingeniería en computadores, que colaboraron con las tareas de mejor forma que los estudiantes de electrónica, pero su disponibilidad horaria fue muy limitada como para permitir el avance necesario.

Además, para el segundo año se contó con fondos menores al año anterior, lo que empeora la situación (se reduce aún más la posibilidad de alcanzar los objetivos), pues no fue posible contratar estudiantes de licenciatura que pudiesen colaborar con más tesis. Otro problema es que para la etapa final del proyecto los temas disponibles se alejaban de los objetivos de un plan de estudios de electrónica, la carrera de computación no tiene estudiantes de licenciatura y hasta finales de 2013 habrá estudiantes de ingeniería en computadores que podrían realizar tesis.

El avance de este proyecto hacia un producto se debe congelar por las razones mencionadas, y se reactivará cuando suficientes estudiantes de maestría y doctorado puedan colaborar con su desarrollo. Esta es la razón por la cual no se solicitó prórroga.

En cuanto a la logística, una limitación importante han sido los lentos procesos burocráticos. En mayo 2010 se solicitó el kit de análisis de suelos, que lamentablemente la empresa envió mal y por tanto tuvo que devolverse hasta obtener el kit solicitado. A pesar de las expresas solicitudes, fue hasta el primero de diciembre que se liberó el monto reservado para ese kit, que por la asignación excesiva de presupuesto en el sistema prácticamente bloqueó un año las compras adicionales. Esto retrasa el avance del

proyecto.

No comprendemos la necesidad de subdivisiones tan excesivas en las entradas del sistema de compras, que es algo únicamente interno a la institución. Es imposible pretender que en procesos de investigación todo ítem posible con todo rango de precios se encuentre en el sistema, y la VIE debería buscar simplificar dicho proceso.



# Capítulo 4

## Observaciones generales y recomendaciones

Se considera que el proyecto ha sido exitoso. Algunos resultados requieren desarrollo adicional para alcanzar nivel de publicación de revista indexada, para lo que se requerirán estudiantes de nivel de maestría o doctorado. En particular las tesis de Randall Esquivel, galardonada con el primer premio del Premio Asoelectrónica 2011 a los mejores Proyectos de Graduación de Ingeniería Electrónica puede elevarse a un artículo de revista, pero se requiere una extensa labor de programación computacional de otros métodos ya expuestos en la literatura para poder comparar resultados. Asimismo, el método propuesto por Pedro Alpízar de difusión controlada en orientación puede elevarse a un artículo, pero también requiere extenso trabajo de programación de otros métodos expuestos en la literatura para poder comparar resultados. Dos borradores de artículo están en proceso, pero por la falta de tiempo de los estudiantes que ya trabajan en la industria y de los investigadores del proyecto que tienen otros compromisos, la programación mencionada avanza lentamente.

A pesar de lo anterior, se publicaron los artículos [2, 11, 8] en un congreso local con asistencia internacional.

Este proyecto ha representado una ventana de colaboración entre profesionales de las áreas biotecnológica, computacional y electrónica, que permitirá futuros proyectos cuando las limitaciones mencionadas sean subsanadas.

Finalmente, el proyecto involucró a siete estudiantes con sus tesis de licenciatura (una de ellas lamentablemente inconclusa) y a diez estudiantes asistentes adicionales. Su contacto con la investigación llevó a varios de ellos a iniciar programas de maestrías académicas.



# Bibliografía

- [1] Antonio Aguilar Bravo. Detección y corrección del efecto sonrisa en imágenes de geles de electroforesis utilizando modelos activos de forma acoplados. Tesis de licenciatura, Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica, Cartago, Abril 2010.
- [2] Pedro Alpízar Salas y Pablo Alvarado-Moya. Anisotropic Diffusion on Electrophoresis Images. In *Proceedings of the Conference on Technologies for Sustainable Development TSD2011*, Cartago, Costa Rica, 2011. URL [http://www.ie.itcr.ac.cr/palvarado/papers/TSD2011\\_Paper\\_06.pdf](http://www.ie.itcr.ac.cr/palvarado/papers/TSD2011_Paper_06.pdf).
- [3] Pedro Elías Alpízar Salas. Optimización de la corrección del efecto sonrisa en imágenes de geles de electroforesis. Tesis de licenciatura, Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica, Cartago, Diciembre 2010.
- [4] P. Alvarado y A. Salazar. Análisis automatizado de patrones de ADN para la caracterización molecular. Informe final. Actividad de fortalecimiento de la investigación, Instituto Tecnológico de Costa Rica, Marzo 2008.
- [5] P. Alvarado, A. Salazar, O. Murillo, F. Rojas, y J. Peraza. Análisis por computador de imágenes de geles de electroforesis para la caracterización molecular de organismos. Informe final, Instituto Tecnológico de Costa Rica, Abril 2010.
- [6] Bryant Esteban Álvarez Canales. Mejoramiento de contraste y razón señal a ruido de imágenes digitales de geles de electroforesis por medio de fusión y ajuste multi-paramétrico de la captura. Tesis de licenciatura, Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica, Cartago, Noviembre 2011.
- [7] Pablo Barrantes-Chaves. Detección de carriles y rectificación de imágenes de geles de electroforesis utilizando modelos activos de forma. Tesis de licenciatura, Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica, Cartago, Junio 2010.
- [8] Pablo Barrantes-Chaves y Pablo Alvarado-Moya. Lane Detection on Gel Electrophoresis Images using Active Shape Models. In *Proceedings of the Conference on Technologies for Sustainable Development TSD2011*, Cartago, Costa Rica, 2011. URL [http://www.ie.itcr.ac.cr/palvarado/papers/TSD2011\\_Paper\\_03.pdf](http://www.ie.itcr.ac.cr/palvarado/papers/TSD2011_Paper_03.pdf).

- [9] Randall José Esquivel Alvarado. Optimización de la detección de cantidad, ancho y posición de bandas en imágenes de geles de electroforesis. Tesis de licenciatura, Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica, Cartago, Junio 2011.
- [10] Edison Fernández Alvarado. Estimación de la desviación estándar para un modelo gaussiano del perfil de bandas en imágenes de geles de electroforesis utilizando información de múltiples escalas. Tesis de licenciatura, Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica, Cartago, Noviembre 2010.
- [11] David Soto-Vásquez y Pablo Alvarado-Moya. Automatic detection of bands in electrophoresis gel images by means of optimization of a target function. In *Proceedings of the Conference on Technologies for Sustainable Development TSD2011*, Cartago, Costa Rica, 2011. URL [http://www.ie.itcr.ac.cr/palvarado/papers/TSD2011\\_Paper\\_22.pdf](http://www.ie.itcr.ac.cr/palvarado/papers/TSD2011_Paper_22.pdf).
- [12] David Soto Vásquez. Detección automática de bandas en imágenes de geles de electroforesis por medio de la optimización de una función objetivo. Tesis de licenciatura, Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica, Cartago, Junio 2010.