

Extracción y recuperación de contenidos basado en vídeo en cirugía de mínima invasión

M.G. Cortina-Januchs^{1,2}, P. Sánchez-González^{1,2}, I. Oropesa^{1,2}, J.A. Sánchez Margallo³,
F.M. Sánchez Margallo³, E.J. Gómez^{1,2}

¹ Grupo de Bioingeniería y Telemedicina, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, España, {mcartina, psanchez, ioropesa, egomez}@gbt.tfo.upm.es

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Madrid, España

³ Centro de Cirugía de Mínima Invasión Jesús Usón, Cáceres, España, {jasanchez, msanchez}@ccmijesususon.com

Resumen

La introducción de las cirugías de mínima invasión en rutina clínica ha provocado la incorporación de los sistemas de vídeo dentro del quirófano. Así, estas técnicas proporcionan al cirujano imágenes que antes solo podían ser vistas mediante cirugía abierta. Los vídeos obtenidos en las intervenciones son almacenados en repositorios. El uso posterior de estos vídeos se ve limitado generalmente a su reproducción, debido a las dificultades de clasificación y gestión. La información que contienen estos repositorios puede ser explotada, reutilizando el conocimiento obtenido en cirugías similares. En este artículo de investigación se presenta el diseño de un módulo de gestión de conocimiento (MGC) para un repositorio de vídeos de cirugía de mínima invasión (CMI). El objetivo del módulo es gestionar y reutilizar la información contenida en el repositorio de vídeos laparoscópicos, para que puedan ser utilizadas con las experiencias previas en entornos de formación de nuevos cirujanos. Para este fin, se han implementado técnicas de recuperación de imagen y vídeo basadas en sus contenidos visuales (CBIR y CBVR). El MGC permite la recuperación de imágenes/vídeos, proporcionando información sobre la tarea que se está realizando en la escena quirúrgica. Los resultados obtenidos en este trabajo muestran la posibilidad de recuperar vídeos de CMI, a partir del instrumental presente en la escena quirúrgica.

1. Introducción

La introducción del vídeo en el quirófano ha supuesto un pilar fundamental para el desarrollo de la cirugía de mínima invasión (CMI). Estas técnicas permiten, entre otras ventajas, disminuir los tiempos de recuperación del paciente [1].

Gran parte de los vídeos grabados en las intervenciones son registrados y almacenados en los repositorios de los hospitales. La información contenida en los repositorios de vídeos de CMI podría ser un recurso de utilidad en el enriquecimiento de historia clínica y en los procesos de formación de cirujanos. A pesar de ello, el uso de estos vídeos se ha visto limitado a su mera reproducción, debido principalmente a la dificultad que implica su clasificación y gestión [2]. Tradicionalmente la edición y etiquetado de los vídeos en los repositorios se lleva a cabo de manera manual mediante palabras clave para su posterior consulta. Al aumentar el tamaño de los repositorios, los métodos manuales de etiquetado se han

vuelto una tarea tediosa y costosa. Por otra parte, esta práctica requiere de un especialista, que sea capaz de describir con palabras clave el contenido de cada vídeo. Además, el idioma y la subjetividad del especialista también pueden plantear problemas cuando se desea consultar algún vídeo en específico [3].

En un intento de solventar la problemática causada por los métodos manuales y facilitar las consultas de los repositorios, la comunidad científica se ha interesado en las técnicas de recuperación multimedia basada en contenidos (CBMR-del inglés Content-Based Multimedia Retrieval), que son útiles cuando las anotaciones de texto son inexistentes o incompletas [2]. Estas técnicas permiten recuperar imágenes/vídeo de manera automática, analizando sus contenidos en lugar de realizar anotaciones manuales. El término contenido se refiere a colores, forma, textura, objetos o cualquier información que se derive de la propia imagen. Además, este tipo de recuperación permite eliminar la subjetividad humana que implica la recuperación por texto [4]. Las aplicaciones médicas se han centrado en su mayoría en técnicas de recuperación de imágenes basada en contenidos (CBIR-del inglés Content-Based Image Retrieval), utilizándose para recuperar estudios de rayos X, tomografía computarizada, ultrasonido principalmente en dermatología y oftalmología [5].

La recuperación de vídeo basado en su contenido (CBVR-del inglés Content-Based Video Retrieval), es una extensión natural de CBIR. La CBVR requiere el análisis de la secuencia de vídeo para extraer los fotogramas representativos, que permitan describir el contenido visual del mismo. Su uso en aplicaciones médicas está menos extendido; en la bibliografía se pueden encontrar aplicaciones en cirugías de retina [6], identificación de flujos en trabajo [7], y clasificación de los procedimientos en CMI [8].

En este trabajo de investigación se presenta un módulo de gestión de conocimiento (MGC) que permite la reutilización de los vídeos de CMI. El módulo permite reconocer la tarea quirúrgica que se está llevando a cabo en base a la detección de las herramientas en la imagen. Para la implementación del módulo se plantea el uso de técnicas de CBVR, CBIR y redes neuronales artificiales.

2. Materiales y métodos

2.1. Materiales

Para el desarrollo de este trabajo se han utilizado secuencias de vídeo quirúrgicas durante el desarrollo de una funduplicatura de Nissen con abordaje laparoscópico usando un modelo porcino. Para ello, se colocaron marcadores biocompatibles de diferente color (verde, azul, naranja y blanco) en los diferentes instrumentales utilizados durante la intervención. Las secuencias fueron adquiridas en el Centro de Cirugía de Mínima Invasión Jesús Usón (CCMIJU) de Cáceres.

2.2. Métodos

El MGC diseñado está compuesto de dos partes: (1) un sistema de recuperación de contenidos (SRC) y (2) un clasificador. En la figura 1 se muestra el diagrama del MGC.

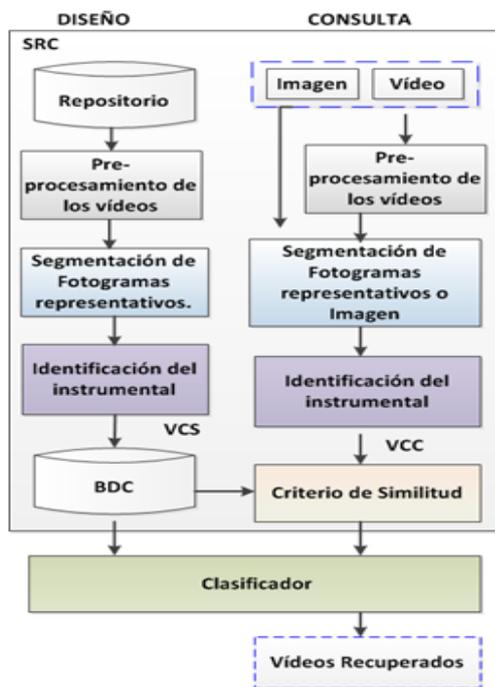


Figura 1. Diagrama del MGC.

El objetivo del SRC es recuperar del repositorio los vídeos que contengan el instrumental presente en la imagen/vídeo de consulta. El clasificador por su parte se encarga de dar una etiqueta a las tareas quirúrgicas que se está realizando en cada imagen/vídeo de consulta. A continuación se detalla cada bloque funcional.

2.2.1. Pre-procesamiento del vídeo

El pre-procesamiento del vídeo se realiza en tres etapas:

1. Eliminación de píxeles de borde. La información de interés en los vídeos endoscópicos se limita típicamente por una región circular o elíptica en el centro de la imagen. Cuando se utiliza la función de zoom, el tamaño del círculo se modifica y el área de interés se vuelve mayor. Además, esta región suele estar rodeada por un borde ruidoso que puede modificar el resultado del procesamiento posterior.

Por esta razón es necesario identificar las regiones que interesa analizar en cada momento. Con la finalidad de determinar el área de análisis se crea una máscara binaria que permita ajustarse al área de interés y eliminar los bordes ruidosos [4].

2. Segmentación del vídeo. Consiste en dividir el vídeo en unidades más pequeñas denominadas segmentos. Estos segmentos permiten identificar un conjunto de fotogramas contiguos, entre los cuales se seleccionarán uno o varios fotogramas representativos. Para la selección de los fotogramas representativos es necesario detectar cambios abruptos entre fotogramas sucesivos, para lo que se calculan las diferencias de los histogramas con la ecuación (1).

$$X^2 = \sum [(H_i - H_{i-1})^2 / (H_i + H_{i-1})^2] \quad (1)$$

Donde H_i es el fotograma actual y H_{i-1} el anterior. Mediante el uso del modelo Dugad (ec. 2) se calcula el umbral m_T para encontrar dichos cambios [9].

$$m_T = \max(\mu_{izq} + T_d (\sigma_{izq})^{1/2}, \mu_{der} + T_d (\sigma_{der})^{1/2}) \quad (2)$$

Donde μ es la media de la ventana deslizante de fotogramas V , σ es la desviación estándar y T_d es una variable calculada empíricamente. T_d determina el rango de valores que serán tomados en cuenta para calcular el umbral m_T .

3. Selección de fotogramas relevantes (FR). Los fotogramas representativos obtenidos de la etapa anterior se analizan y se eliminan aquellos fotogramas irrelevantes (FI). Las imágenes que muestran el interior del paciente se caracterizan por presentar colores rojizos o amarillentos, debido a que contienen tejidos y sangre. Por su parte las imágenes de la sala de operaciones (exterior del paciente), suelen presentar colores verdes o azules, típicamente utilizados en la ropa de los cirujanos. Se utiliza el espacio de color HSV (por sus siglas en inglés, Hue-Saturation-Value) para discriminar entre los FR y los FI. Se selecciona un conjunto de fotogramas que representen FR y FI y se calculan los histogramas de H y S . Los histogramas obtenidos son agrupados mediante el algoritmo Kmeans [10] en 2 grupos que representan los FR y FI.

2.2.2. Segmentación de los fotogramas representativos

El bloque de segmentación del fotograma tiene dos funciones principales:

1. Eliminación de brillos de cada imagen. El objetivo de la eliminación de brillos es reducir los efectos de la iluminación no uniforme típica de las imágenes laparoscópicas. Con ese fin se implementa una apertura por reconstrucción de erosión en cada una de las componentes de color de la imagen.
2. Segmentación de los marcadores de color. La segmentación de los fotogramas representativos se realiza buscando las regiones homogéneas de los marcadores de color colocados en los instrumentos quirúrgicos. Cada FR es segmentado utilizando el espacio de color HSV.

2.2.3. Identificación del instrumental

Con las imágenes binarias obtenidas de la segmentación se crean los vectores de instrumental VCs (en el diseño) y el VCc (en la consulta). Cada marcador de color se corresponde una herramienta en la escena quirúrgica. Los vectores indican qué instrumental está presente en los FR de cada vídeo y la orientación de cada instrumento. Los VCs son almacenados en la base de datos de características (BDC) para posteriormente ser utilizados durante la consulta.

2.2.4. Criterio de similitud

Uno de los mayores retos en los sistemas de recuperación de contenidos es definir la medida adecuada de similitud. Las medidas de similitud usadas en estos sistemas suelen ser las distancias Euclídea, Manhattan, Chevychev y Minkowski [11]. Otros métodos miden la similitud en términos de información de la imagen, usando medidas de similitud basadas en la clasificación de las imágenes [12]. En este trabajo se utiliza una combinación de la distancia Euclídea y un clasificador (explicado en el siguiente apartado) como criterio de similitud. Se calcula la distancia del VCc con los VCs del repositorio, obteniendo los vídeos más similares.

2.2.5. Clasificador

El clasificador tiene la función de asignar una etiqueta a cada fotograma de los vídeos dependiendo de la tarea quirúrgica (duplicatura, disección, manipulación, etc.) que se está realizando. El diseño del clasificador se realiza en dos etapas: (1) agrupamiento de los VCs y (2) clasificación de las tareas que se realizan en la escena quirúrgica basado en una red neuronal artificial (RNA).

En la primera etapa, se agrupan los VCs para encontrar las combinaciones existentes entre los distintos instrumentos, asignando etiquetas a cada grupo [12]. Una vez agrupados los VCs, en la segunda etapa se diseña una RNA. La RNA aprenderá las etiquetas que fueron asignadas durante el agrupamiento [13]. En este trabajo se ha diseñado una RNA perceptrón multicapa con 16 neuronas en la capa de entrada y 5 en la capa de salida. La salida de la RNA nos proporciona una inferencia del tipo de tarea que se está realizando de acuerdo a la combinación del instrumental presente en la imagen.

2.3. Evaluación

El rendimiento del sistema se realiza evaluando los resultados obtenidos en las consultas (imágenes/vídeos). Los vídeos utilizados fueron divididos en 15 segmentos y etiquetados, ambos procesos automáticamente. Las métricas utilizadas para evaluar el rendimiento del MGC son la precisión y exhaustividad (ec. 4 y 5).

$$\text{Precisión} = \text{VRR} / \text{VR} \quad (4)$$

$$\text{Exhaustividad} = \text{VRR} / \text{VRE} \quad (5)$$

Donde VRR son el número vídeos relevantes recuperado, VR los videos recuperados, y VRE los vídeos relevantes en el repositorio.

3. Resultados y discusión

Los resultados obtenidos muestran que el modelo Dugad permite identificar los segmentos de los vídeos de CMI automáticamente. Cada segmento representa acciones realizadas durante la cirugía, por ejemplo: introducción del laparoscopio, manipulación del hígado, introducción del instrumental, etc. En la figura 2 se muestran algunos ejemplos de los fotogramas seleccionados en la segmentación del un vídeo, con ventana deslizante $V=500$ y $T_d=2$. El modelo Dugad calcula el umbral en cada ventana deslizante. De este modo es posible encontrar cambios muy grandes, como en las imágenes 1, 3 y 4. El tamaño de V , permite encontrar también cambios más pequeños pero significativos en X^2 , como el que se muestra en la imagen 2, donde se observa que un instrumental verde es introducido. Un valor pequeño de la distancia X^2 en cada ventana representa pocos cambios entre los fotogramas contiguos.

El modelo Dugad permite reducir a un 2,3% los fotogramas seleccionados de cada vídeo. La elección de V y T_d afecta el tamaño de los segmentos en los que se divide cada vídeo; cuanto mayor sean los valores de V y T_d , mayor será el tamaño de los segmentos y menor el número de fotogramas seleccionados. Si bien esto puede ser una ventaja en términos computacionales, una elección incorrecta puede generar segmentos grandes que no representen la información del instrumental en la escena quirúrgica.

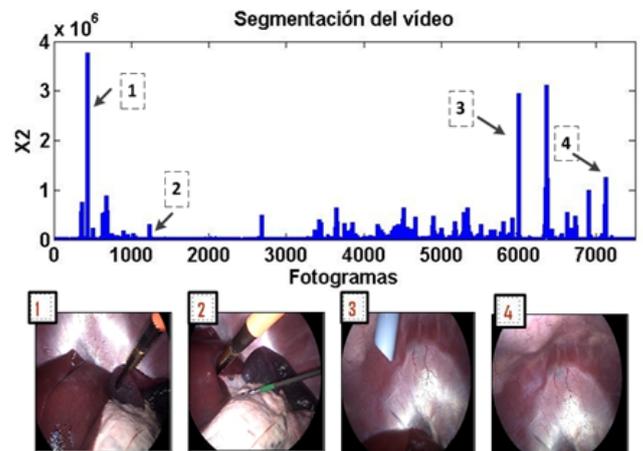


Figura 2. Segmentación de un vídeo con $V=500$ y $T_d=2$. Se muestran 4 ejemplos de FR, donde se observa que el umbral m_r , varía considerablemente dependiendo de los valores de X^2 en cada una de las ventanas deslizantes.

Una vez obtenidos los FR de cada vídeo, se realiza la segmentación de los marcadores de color para identificar el instrumental presente y formar los vectores de instrumental (VCc). La figura 3 muestra la segmentación obtenida para dos imágenes.

Los resultados obtenidos de la segmentación de los FR muestran que los marcadores de color son de utilidad para identificar el instrumental. Se observa que los marcadores de color azul, verde y naranja, pueden ser segmentados con mayor facilidad en la escena quirúrgica. Por otro lado, el marcador color blanco puede confundirse con los brillos originados por la incidencia de la luz con algunos

tejidos, provocando errores en la segmentación y en los vectores de instrumental.

Una vez creados los vectores de instrumental, se clasifica la tarea que se está realizando en la imagen/vídeo. En la figura 3, se muestran dos imágenes de distintas tareas quirúrgicas, a cada una el clasificador asigna una etiqueta de acuerdo al instrumental presente, en este caso las etiquetas asignadas son: la duplicatura y la disección.

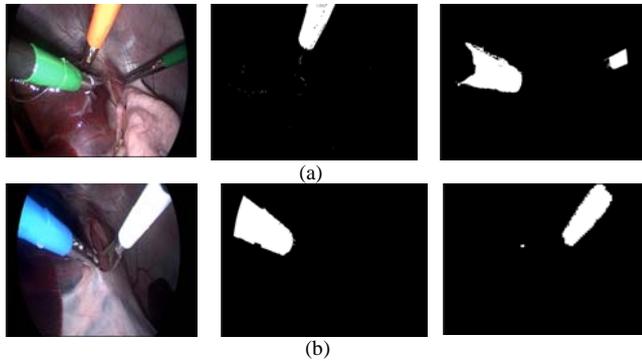


Figura 3. Segmentación de FR: (a) duplicatura, donde intervienen tres instrumentos: pinza de tracción (verde-izquierda), disector (verde-derecha) y porta agujas (naranja) (b) disección, donde intervienen dos instrumentales: pinza de tracción fenestrada (azul-izquierda) y tijeras (blanco-derecha).

El MGC fue evaluado con 10 imágenes y 10 vídeos de búsqueda. Para las imágenes evaluadas se obtuvo una precisión mínima de 0,7 y una precisión máxima de 1. En las búsquedas por vídeos se obtuvo una precisión mínima de 0,75 y una precisión máxima de 1. El promedio de las búsquedas realizadas se muestra en la tabla 1.

Consulta	Precisión	Exhaustividad
Imágenes	0,84	0,87
Vídeos	0,9	0,84

Tabla 1. Resultados obtenidos del MGC.

Los resultados de las búsquedas realizadas en el MGC muestran que es posible la recuperación de vídeos de CMI, mediante técnicas de segmentado de vídeo y técnicas de recuperación de contenidos.

4. Conclusiones

El manejo base de datos multimedia de gran tamaño, requiere de técnicas adecuadas para su recuperación y exploración. En particular, los repositorios de vídeos de CMI requieren de sistemas que recuperen información de manera eficiente y lo más parecidas a la imagen/vídeo de consulta, para proporcionar información que pueda ser relevante al usuario. La extracción y recuperación eficiente del conocimiento contenido en los repositorios de vídeos de CMI pueden aportar información que ayude a la decisión clínica, los procesos de formación de nuevos cirujanos y la formación continuada de los profesionales.

Los resultados preliminares obtenidos en este trabajo muestran que es posible recuperar vídeos laparoscópicos, a partir del instrumental presente en la escena quirúrgica.

Las técnicas implementadas ofrecen buenos resultados para los vídeos con marcadores de color verde, azul y naranja. Para la completa validación del MGC es necesaria la introducción de más vídeos laparoscópicos y con diferentes contextos quirúrgicos.

Agradecimientos

Los autores quieren agradecer la financiación del proyecto de investigación al Consejo Nacional de Ciencia y Tecnología de México (CONACyT) y los medios técnicos y personales prestados por el Grupo de Bioingeniería y Telemedicina de la ETSI de Telecomunicación y el Centro de Cirugía de Mínima Invasión Jesús Usón.

Referencias

- [1] Cuschieri A. Laparoscopic surgery: Current status, issues and future developments, *The Surgeon*, vol.3, no.3, pp. 125-138, 2005, (ISSN: 1479-666X)
- [2] Marques O, Furht B. Content based image and video retrieval. Vol.1. Springer press, 2002 (ISBN:978-1-4613-5338-6)
- [3] Lew MS, Sebe N, Djeraba C, Jain R. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Computing Communication and Application*, vol.2, pp.1-19, 2006, (ISSN:1551-6857)
- [4] Munzer B, Schoeffmann K, Boszormenyi L, Relevance Segmentation of Laparoscopic Videos. *IEEE International Symposium on Multimedia (ISM'13)*, California, USA 2013, pp. 84-91 (ISBN: 978-0-7695-5140-1)
- [5] Akbarpour Sh. A review on content based image retrieval in medical diagnosis. *International journal on technical and physical problems of engineering*, 2013, vol.5, no.2, pp. 158-153
- [6] Droueche Z, Lamard M, Cazuguel G, Quellec G, Roux C, Cochener B. Motion-based video retrieval with application to computer-assisted retinal surgery. *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp.4962-4965, Aug. 28 2012-Sept. 1 2012
- [7] Giannarou S, Yang GZ. Content-based surgical workflow representation using probabilistic motion modeling, in *LNCSS Medical Imaging and Augmented Reality*, vol. 6326, 2010, pp. 314-323
- [8] Blum T, Feubner H, Navab N. Modeling and segmentation of surgical workflow from laparoscopic video. *Medical image computing and computer-assisted intervention*, 2010, vol. 6363, pp.400-407
- [9] Krulikovská L, Polec J. Shot Detection using Modified Dugad Model. *World Academy of Science, Engineering and Technology*, vol.6, pp.117-120, 2012 (ISSN:1307-6892)
- [10] MacQueen JB. Some methods for classification and Analysis of multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281-297 (ISBN:0-520-02188-6)
- [11] Darsana B, Jagajothi G. DICOM Image Retrieval Based on Neural Network Classification, *International Journal of Computer Science and Telecommunications*, vol.5, no.3, pp. 21-26, 2014
- [12] Oberoi, A, Bakshi V, Sharma R, Singh M. A Framework for Medical Image Retrieval Using Local Tetra Patterns. *International Journal of Engineering and Technology*, vol. 5, no.1, pp.27-36, 2013.
- [13] Haykin S. *Neural Networks: A Comprehensive Foundation*. Prentice hall press, 1998 (ISBN:0132733501)