**IMPROVED DCT COEFFICIENT ANALYSIS FOR FORGERY LOCALIZATION IN JPEG IMAGES**

**INTRODUCCION**:

La gran cantidad de imágenes digitales que se extienden por todas partes y la facilidad de uso de las herramientas de procesamiento de imágenes comerciales, hacen que la difusión de contenidos falsos a través de la red digital sea cada vez mayor y preocupante. Una imagen resultante del empalme de diferentes fotos transmite una información distante del evento original capturado por la fuente original: la información transmitida por la imagen manipulada es falsa e influye en la opinión de los espectadores que creen la credibilidad de la imagen que están viendo.

Este problema tiene efectos importantes en varios ámbitos en los que debe comprobarse la credibilidad de las imágenes antes de utilizarlas como fuentes de información, como por ejemplo, el seguro, la ley y el orden, el periodismo y las aplicaciones médicas.

Durante los últimos años se han propuesto muchos algoritmos forenses de imágenes para exponer falsificaciones digitales por medio de diferentes enfoques basados ​​en el análisis de características estadísticas y geométricas, artefactos de cuantificación JPEG, efectos de interpolación, rastros de demostración, inconsistencias de características, etc. [ 1]. Incluso pensando que tales herramientas forenses parecen ser útiles para la verificación de la integridad de la imagen, por otro lado, su aplicación en escenarios reales aún no ha sido validada adecuadamente. De hecho, la mayoría de los enfoques propuestos hasta ahora buscan la presencia/ausencia de algunas características dentro de una región dada y se basan en la hipótesis para conocer la ubicación correcta del área de falsificación. Esta cuestión debe ser cuidadosamente considerada cuando las técnicas forenses se aplican en casos prácticos, donde las imágenes se proporcionan sin ninguna información a priori.

Dado que una gran cantidad de imágenes de hoy se comprime en formato JPEG, varias herramientas forenses han explorado las características de los coeficientes DCT para revelar cierta incoherencia dentro de la imagen [2][3], pero incluso en este caso la detección de manipulación local sigue siendo una cuestión abierta. Se ha hecho algún intento en tal dirección, por ejemplo, en [4] autores proponen un algoritmo para la detección de regiones de imagen que se han trasplantado de otra imagen y en [5] autores presentaron un método para la localización automática de aleatoriamente regiones en imágenes JPEG, con una escala de grano fino de 8 × 8 bloques.

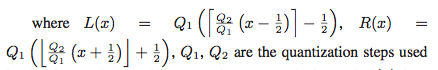
El objetivo de este trabajo es avanzar un paso más en esta dirección, proponiendo una versión mejorada de [5], donde los autores calculan un mapa de probabilidad (descrito en la Sección 2) para diferenciar regiones alteradas y no modificadas. El modelo utilizado para construir el mapa tiene algunas limitaciones principales que serán profundamente analizados y superados por el nuevo enfoque (Sección 3), lo que conduce a una mejora significativa de la precisión de la estimación de mapa de probabilidad y, en consecuencia, del rendimiento del algoritmo, como se muestra En la Sección 4.

**PREVIOUS WORK:**

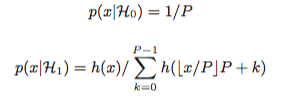
En muchos escenarios, una imagen JPEG original, después de una falsificación localizada, se guarda de nuevo en formato JPEG sin cambiar el tamaño. Por lo tanto, los coeficientes DCT de áreas no modificadas sufrirán una doble compresión JPEG, exhibiendo así artefactos de doble cuantificación (DQ), mientras que los coeficientes DCT de áreas forjadas resultarán de una compresión única y probablemente no presentarán artefactos. A continuación, nos referiremos a este escenario como la única hipótesis de falsificación de compresión (SCF). En [5], se propone en primer lugar utilizar las estadísticas de los coeficientes DCT de las imágenes JPEG doblemente comprimidas para discriminar entre áreas manipuladas y originales bajo la hipótesis del SCF. La idea es utilizar la inferencia bayesiana para asignar a cada coeficiente DCT una probabilidad de ser doblemente cuantificada. Dichas probabilidades, acumuladas sobre cada bloque 8 × 8, proporcionarán un mapa de probabilidad DQ que nos permitirá indicar las áreas originales (alta probabilidad de DQ) de las áreas alteradas (baja probabilidad de DQ).

La inferencia bayesiana se basa en la distribución de probabilidad de los coeficientes DCT condicionales a la hipótesis de ser manipulado, es decir, p (x | Hb), donde x es el valor del coeficiente DCT y H0 (H1) indica la hipótesis de ser manipulado. En [5], tales probabilidades condicionales se derivan observando que el histograma de los coeficientes DCT que tienen la misma posición dentro de un bloque 8 × 8 exhibe un patrón periódico después de la cuantificación doble. Concretamente, el número de compartimientos del histograma original que se asignan en el contenedor correspondiente al valor x en un histograma DQ están dados por una fórmula:

../../../../Desktop/Captura%20de%20pantalla%202017-04-06%20a%20las%2015.29.2



donde Q1, Q2 son las etapas de cuantificación utilizadas en la compresión primera y segunda, respectivamente, y n(x) es una función periódica, con periodo P = Q1 / gcd (Q1, Q2). Una vez que P se ha estimado a partir del histograma de los coeficientes DCT (ver [5] para más detalles) los autores de [5] proponen estimar las probabilidades condicionales como



Donde h es el histograma de los coeficientes de DCT1. Bajo la hipótesis de que los coeficientes DCT dentro de un bloque son mutuamente independientes y que H0, H1 son equiprobables, la probabilidad de que un bloque sea alterado puede estimarse como

../../../../Desktop/Captura%20de%20pantalla%202017-04-06%20a%20las%2015.35.2

En [5], algunas características extraídas de la probabilidad mapa dado por (4) son añadidos a un clasificador con el fin de detectar automáticamente si la imagen es alterada o no.

**ANÁLISIS DEL COEFICIENTE DCT PROPUESTO:**

Una limitación del método descrito anteriormente es que la probabilidad condicional p (x | H1) se estima de acuerdo con el histograma observado de x. En el caso de una imagen manipulada, sin embargo, tal histograma será realmente una mezcla de p(x|H1) y p(x|H0). Por lo tanto, para falsificaciones grandes, esperamos que el histograma de x sea una estimación pobre de p(x|H1).

Para superar esta limitación, deberíamos ser capaces de separar las dos probabilidades condicionales de la mezcla observada. Suponiendo que el histograma h0(x) de los coeficientes DCT antes de la primera compresión JPEG está disponible, una mejor estimación de p(x|H1) podría obtenerse como

**../../../../Desktop/Captura%20de%20pantalla%202017-04-06%20a%20las%2015.38.4**

Desafortunadamente, (5) es difícil de usar en la práctica, ya que requieren una estimación fiable tanto de h0(x) como de Q1. Por lo tanto, proponemos introducir la siguiente aproximación

../../../../Desktop/Captura%20de%20pantalla%202017-04-06%20a%20las%2015.41.3

Donde L’(x) = Q2x – Q2/2 y R’(x) = Q2x + Q2/2. La aproximación se cumple cuando n(x)> 0 y el histograma del coeficiente de DCT original es localmente uniforme. En la práctica, se encontró que para valores moderados de Q2 esto es generalmente cierto, excepto para el bin central (x = 0) de los coeficientes AC, que tienen una distribución similar al Laplaciano. Obsérvese que el lado derecho de (6) se obtiene remuestreando h0(x) con la etapa Q2, es decir, h(x) puede ser visto como el histograma de los coeficientes de DCT después de una sola compresión con la etapa de cuantificación Q2. Una técnica simple para estimar h(x) es considerar los coeficientes DCT obtenidos mediante la recompresión con la segunda matriz de cuantificación de una versión ligeramente recortada de la imagen falsificada [6].

De acuerdo con la hipótesis del SCF, podemos estimar las probabilidades condicionales como

../../../../Desktop/Captura%20de%20pantalla%202017-04-06%20a%20las%2015.46.4

../../../../Desktop/Captura%20de%20pantalla%202017-04-06%20a%20las%2015.47.0

Basándose en las ecuaciones anteriores, la probabilidad de que un bloque sea manipulado puede ser simplemente estimada como

../../../../Desktop/Captura%20de%20pantalla%202017-04-06%20a%20las%2015.47.1

Donde ni(xi) indica la función n(x) relacionada con la i-ésima DCT coeficiente dentro de un bloque. Tenga en cuenta que dado que tanto (7) como (8) pueden no ser exactos cuando x = 0, el cálculo real del mapa de probabilidades no tiene en cuenta los coeficientes de DCT igual a cero. Además, debido al hecho de que la mayor parte de los coeficientes de DCT son cero a altas frecuencias, en la práctica sólo se usan los primeros coeficientes de baja frecuencia.

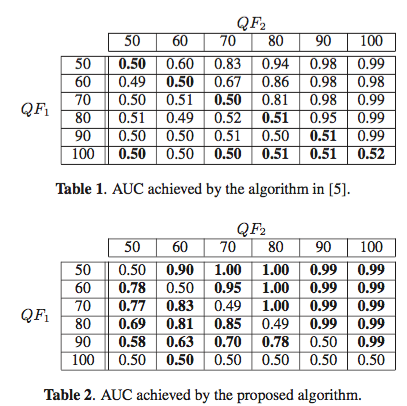
**ANÁLISIS DE RENDIMIENTO**:

En esta sección se describe en primer lugar la metodología experimental que seguimos para evaluar el desempeño del detector de falsificación mejorado y para proporcionar una comparación adecuada con el anterior. Entonces, mostramos los resultados experimentales que vienen de tal análisis.

El conjunto de datos de imagen utilizado para probar los algoritmos está compuesto por 100 imágenes TIFF no comprimidas, con un contenido heterogéneo, procedentes de tres cámaras digitales diferentes (concretamente, Nikon D90, Canon EOS 450D y Canon EOS 5D). La porción central de tamaño 1024 × 1024 se extrae de cada imagen para formar el conjunto de datos original. A partir de esto, creamos las correspondientes imágenes manipuladas siguiendo la hipótesis del SCF. Es decir, cada imagen original es comprimida en JPEG con un factor de calidad QF1 dado (usando la función Matlab imwrite). La parte central de tamaño 256 x 256 se reemplaza entonces con el área correspondiente de la imagen TIFF original.

Finalmente, la imagen global "manipulada" es JPEG comprimida (de nuevo con Matlab) con otro factor de calidad dado QF2. De esta manera, la imagen resultará doblemente comprimida en todas partes, excepto en la región central donde se supone que se forja. Tanto QF1 como QF2 se toman del conjunto [50, 60,... 100] obteniendo 36 combinaciones posibles de (QF1, QF2) para cada una de las 100 imágenes manipuladas.

La selección de una métrica de rendimiento adecuada es fundamental en esta etapa. Ambos algoritmos considerados proporcionan como salida, para cada imagen analizada, un mapa de probabilidad que representa la probabilidad de que cada bloque 8 × 8 sea forjado (es decir, para cada imagen 1024 × 1024, se da un mapa de probabilidad de 128 x 128). Después de una etapa de umbral, se logra un mapa de detección binario, que localiza cuáles son los bloques detectados como alterados.



Supongamos ahora que se conoce para cada imagen analizada la posición de las áreas forjadas. Ya que en nuestras pruebas experimentales se modifica una región central 256 × 256 para que se comprima individualmente, es posible asociar a cualquier imagen manipulada una correspondiente máscara binaria 128 × 128 cuya porción central 32 × 32 indica bloques forjados. Una comparación entre el mapa de detección de salida del algoritmo y la máscara de manipulación conocida permitirá estimar las tasas de error de los esquemas forenses, medidos como probabilidad de falsa alarma Pfa y probabilidad de detección perdida Pmd. Estas dos probabilidades se pueden calcular midiendo los siguientes parámetros:

nNMF: bloques numéricos no manipulados, pero detectados como forjados;

NMNF: número de bloques manipulados, pero no detectados como forjados;

NI: número de bloques en la imagen (16384 en nuestras pruebas);

NM: número de bloques manipulados (1024 en nuestras pruebas).

Partiendo de estas cifras, las probabilidades de error son dadas por:

/Users/mariasolana/Desktop/Captura de pantalla 2017-03-31 a las 10.49.38.png

Y la probabilidad de detección correcta es: Pd = 1 – Pmd.

Para representar la compensación entre la tasa de detección correcta Pd y la tasa de falsas alarmas Pfa, se hace referencia a la curva característica de funcionamiento del receptor (ROC) y la consideramos un medio apropiado para medir las prestaciones del detector de falsificación. En particular, dado que la curva ROC es un gráfico bidimensional de Pd frente a Pfa cuando se varía el umbral de decisión del detector, adoptamos el área bajo la curva ROC (AUC) para resumir el rendimiento con un valor escalar único que representa el comportamiento general del detector. Se sabe que el AUC debería asumir valores entre 0,5 y 1 para detectores realistas y efectivos (es decir, no aleatorios).

En la práctica, movimos el umbral de decisión de 0 a 1 con el paso 0.01 y para lograr el AUC correspondiente a cualquier combinación de (QF1, QF2) promediamos los resultados de Pfa y Pd calculados en todas las 100 imágenes alteradas. Los valores de AUC obtenidos por el algoritmo en [5] y por el método mejorado se muestran en las tablas 1 y 2, respectivamente. Los mejores resultados para cada combinación (QF1, QF2) se resaltan en negrita y demuestran que el nuevo mapa de probabilidad tiene una precisión mejorada que ayuda a discriminar regiones forjadas y no cambiadas. En particular, el método propuesto es capaz de revelar adulteraciones incluso si QF2 <QF1, que a menudo se produce en casos prácticos.

El algoritmo también ha sido probado en un conjunto de imágenes que representan casos realistas de falsificación. En la Figura 1 se muestran dos ejemplos de imágenes sospechosas: nuestro mapa de probabilidad revela claramente un área donde una tercera persona podría haber sido removida y una región rectangular que puede haber sido pegada de otra imagen, mientras que los mapas de probabilidad del método anterior sólo proporcionan información fuertemente dependiente del contenido de la imagen.