

DETECCIÓN DE CARACTERÍSTICAS EN IMÁGENES BASADAS EN EL TENSOR DE COLOR

Filadelfio Caballero, José María. Sebastián
Departamento de Automática, Ingeniería Electrónica e Informática Industrial (DISAM).
Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid.
C/ José Gutiérrez Abascal, 2, 28006 Madrid, España.
filadelfio.caballero@utp.ac.pa, jsebas@etsii.upm.es

Resumen

La correspondencia entre dos imágenes sigue siendo un de los aspectos más críticos de la visión por computador. Su resolución se emplea en numerosas tareas, entre las que destacan: Reconocimiento de objetos o escenas, resolver una estructura 3D entre múltiples imágenes, correspondencia estéreo, y seguimiento de objetos. El descriptor Sift ha demostrado ser el mejor detector de características invariantes en imágenes en escala de gris; sin embargo cuando la imagen tiene una pobre iluminación el detector de Sift presenta deficiencias en poder realizar el casamiento en las características detectadas, es decir, es poco robusto a los cambios bruscos de iluminación. Por otro lado la información del color puede ser utilizada para compensar, complementar o reemplazar a un descriptor de forma, dada la gran cantidad de información que se puede extraer del modelo de color.

En este trabajo proponemos un descriptor de color basado en el Tensor de Color de Di Zenzo, el cual mejora el desempeño del Sift original, manteniendo las invariancias con respecto a traslaciones, rotaciones, cambios de iluminación, sombras y reflejos. En general el descriptor de color puede ser aplicado a una amplia variedad de imágenes, donde se detecten y realicen el match de estas características utilizando solo el criterio del color.

Palabras Clave: Tensor de Color, Sift, Gradiente de Color.

1 Introducción

SIFT (Scale Invariant Feature Transform), es un método para extraer características distintivas de las imágenes en escala de grises, es decir que puede ser utilizado para reconocer la misma característica entre diferentes vistas de un mismo objeto o escenas. El método fue diseñado por David G. Lowe [5]. Las características extraídas son invariantes a la escala de la imagen y a la rotación, parcialmente es invariante

a la deformación Afín, cambios en vistas 3D, adición de ruido y cambios en la iluminación.

El algoritmo analiza una imagen a través de la escala de espacio utilizando el método piramidal con sucesivos filtros Gaussianos. Este encuentra máximos y mínimos de contraste a través de tres máscaras 3x3 en tres niveles adyacentes de diferencias Gaussianos (DOG), los cuales son los potenciales puntos claves. Estos puntos claves son verificados para descartar aquellos que son inestables, y para los que resulten estables se genera un descriptor. Los descriptores representan la información local de la imagen alrededor del punto clave de tal forma que es invariante a los cambios en escala, rotación y pequeñas variaciones de iluminación, lo que significa que las mismas características van a ser detectadas en la imagen a un si los objetos han sido rotados o son vistos desde una perspectiva diferente o ha cambiado la distancia.

SIFT es un método que no es totalmente invariante a la transformación Afín, pero si permite desplazamientos relativos de las características de la imagen generando pequeños cambios en el descriptor. SIFT genera un descriptor de histogramas de localización y orientación de gradientes. La localización es cuantizada en un grid o rejilla de 4x4 y el ángulo de gradiente es cuantizado en 8 orientaciones. Resultando un descriptor con dimensión de $4 \times 4 \times 8 = 128$. El método consiste en coleccionar una gran cantidad de características locales (vectores), cada uno de los cuales es invariante a la traslación, escalamiento, y rotación. Parcialmente es invariante a cambios de iluminación y a la proyección 3D o AFIN.

2 Procesamiento de Imágenes a Color

Se han desarrollado diversos métodos para tomar en cuenta el color en la detección de características invariantes, sin embargo algunos de estos métodos se basan en el análisis de los canales de color por separado. Entre ellos podemos mencionar:

- Geodeme, utilizó momentos de color con Sift.

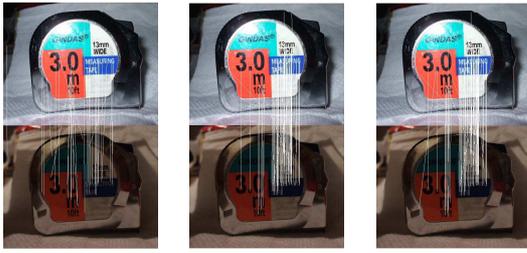


Figura 1: En la imagen izquierda los resultados obtenidos para el match entre los planos S del modelo HSV con 740 características y 35 match points. En la imagen central, el match de características en el modelo CIELab con 355 características detectadas y 89 características en común. En la imagen derecha, el match de características en el modelo CIELuv con 91 características en común.

- Abdel-Hakim[1], utilizó invariantes del modelo de color Gaussiano con Sift.
- A. Bosch[2] utilizó HSV y Sift.
- Burghouts[3], utilizó invariantes fotométricas en los canales del color opoente para determinar el gradiente de color dentro del Sift.
- J. van de Weijer[8], utilizó el modelo del color opoente y sus derivativas invariantes con el Sift.
- P. Schugerl[7], utilizó el descriptor de color MPEG-7 en la extracción de puntos de interés.

Varias aproximaciones hemos desarrollado en el análisis de imágenes a colores que mejoran significativamente el método de Sift original, pero no eran detecciones directas sobre el modelo de color RGB. Entre estas aproximaciones podemos mencionar (1)La detección de características en el espacio de color HSV y el cual es mejorado en los espacios de color CIELab y CIELuv. En estos análisis se comprobó que se pueden detectar más características distintivas en los planos H y S para el espacio HSV y en el plano X para el espacio CIELab y CIELuv ya que estas características permacen constantes después de un filtrado Gaussiano, los resultados se muestran en la figura 1.(2)La detección sobre los planos de color RGB en forma independiente y luego agrupando las diferentes contribuciones de cada plano. En este caso se realiza el procedimiento del Sift original sobre cada plano de intensidad dando como resultado una mayor cantidad de match point, como se muestra en la figura 2 y 3.

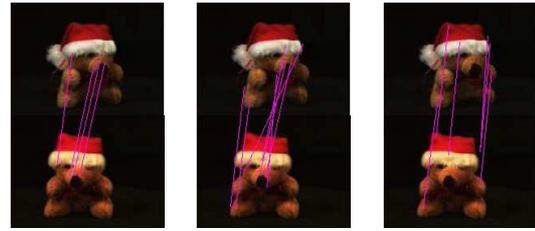


Figura 2: Resultados obtenidos con el sift. Se muestra el match de cada plano de color R, G y B respectivamente.

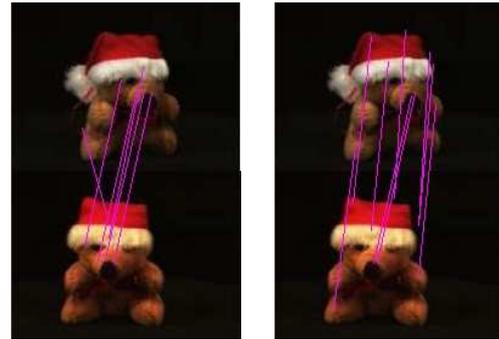


Figura 3: Resultados obtenidos con el sift. A la izquierda se muestra el match obtenido en escala de grises(Sift original). A la derecha se observa el match obtenido como una contribución de todos los planos de color.

El método directo consiste en incorporar todas las componentes de color en la determinación del descriptor Sift. Este método presenta dificultades ya que los canales de color son muy sensitivos a los cambios de iluminación. El método consiste en procesar los datos de los tres planos al mismo tiempo para encontrar la característica tanto en la detección como en la descripción. En la primera fase la imagen original se pasa a un formato de 32 bits y tres canales, esta imagen se dobla, para aumentar el número de keypoints antes del filtrado y luego se suaviza con un filtro gaussiano. La imagen resultante se pasa como la imagen base para la construcción de la pirámide gaussiana. El procedimiento de suavizado se puede realizar de dos formas: directamente sobre la imagen a color o por planos de color separados, con ambos procedimientos se obtuvieron los mismos resultados. Se trabajó sobre la base de imágenes en la que el SIFT original tuviera problemas para hacer un apropiado match de una significativa cantidad de keypoints, como se muestra en la figura 4. El objetivo específico era el de obtener una mayor cantidad de keypoints y lograr hacer una mayor cantidad de match points.



Figura 4: Resultados obtenidos con el sift original en escala de grises. En la imagen izquierda no se ha logrado ningún match y en la derecha pocos matches points.

3 Tensor de Color.

El Tensor de Color es un ente matemático con el cual se puede expresar un concepto multilinear en este caso una matriz de componentes de derivadas de color. El Tensor de Color ha sido utilizado para la segmentación de imágenes a color[6], para la detección de características en imágenes a color[10] y [9]. Este concepto fue desarrollado por Di Zenzo[4] donde establece que la simple suma de las derivadas ignora la correlación entre los canales de colores. Esto también sucede cuando se convierte la imagen de color a valores de luminancia. Debido a que las derivadas de un borde apuntan en direcciones opuestas para canales separados, la simple suma de estructuras diferenciales de varios canales de color puede resultar en la cancelación de esta estructura en la imagen. Más que sumar la dirección de la información de los canales, es más apropiado la suma de la información de la orientación. Tal método es suministrado por un tensor en el que los vectores con direcciones opuestas se refuerzan uno al otro. El tensor describe la orientación local en vez de una dirección. Más precisamente el tensor de un vector y su vector rotado 180° son iguales[8].

Di Zenzo propone que una forma correcta de combinar las derivativas de primer orden es utilizar un Tensor de Color local. Un trabajo extenso sobre el Tensor de Color fue realizado por J. van de Weijer [8]. Un análisis de la forma del Tensor conduce a una orientación y a la norma del gradiente. La estructura del Tensor está dada por la ecuación 1 como:

$$G = \begin{pmatrix} \overline{f_x^2} & \overline{f_x f_y} \\ \overline{f_x f_y} & \overline{f_y^2} \end{pmatrix} \quad (1)$$

Para una imagen a color:

$$f = (R, G, B)^T \quad (2)$$

la estructura del Tensor de color es la ecuación 3:

$$G = \begin{pmatrix} \overline{R_x^2 + G_x^2 + B_x^2} & \overline{R_x R_y + G_x G_y + B_x B_y} \\ \overline{R_x R_y + G_x G_y + B_x B_y} & \overline{R_y^2 + G_y^2 + B_y^2} \end{pmatrix} \quad (3)$$

Donde la barra superior indica una convolución con un filtro Gaussiano. El Tensor de Color describe la estructura diferencial de primer orden en un punto de la imagen 2D, y puede ser descompuesta resolviendo la ecuación característica:

$$G - (\lambda * I)e = 0 \quad (4)$$

donde λ son los valores característicos del tensor y e son los vectores unitarios. Las soluciones no triviales se obtienen si:

$$|G - (\lambda * I)| = 0 \quad (5)$$

$$\begin{vmatrix} \overline{f_x^2} - \lambda & \overline{f_x f_y} \\ \overline{f_x f_y} & \overline{f_y^2} - \lambda \end{vmatrix} = 0 \quad (6)$$

Los valores característicos de la estructura del Tensor están dados por la ecuación 7 y 8:

$$\lambda_1 = \frac{1}{2}(\overline{f_x^2} + \overline{f_y^2} + \sqrt{(\overline{f_x^2} - \overline{f_y^2})^2 + (2\overline{f_x f_y})^2}) \quad (7)$$

$$\lambda_2 = \frac{1}{2}(\overline{f_x^2} + \overline{f_y^2} - \sqrt{(\overline{f_x^2} - \overline{f_y^2})^2 + (2\overline{f_x f_y})^2}) \quad (8)$$

La orientación en la imagen con el máximo cambio de color queda determinado por la ecuación 9:

$$\theta = \frac{1}{2} \arctan \left(\frac{2\overline{f_x f_y}}{\overline{f_x^2} - \overline{f_y^2}} \right) \quad (9)$$

Los valores característicos pueden ser combinados para obtener los siguientes descriptores:

- λ_1 es la energía derivativa en la dirección más prominente.
- λ_2 describe la cantidad de energía derivativa perpendicular a la dirección local prominente que es utilizada para seleccionar características para el seguimiento de objetos.

- $\lambda_1 + \lambda_2$ describe la energía derivativa total, también llamado traza.
- $\lambda_1 - \lambda_2$ describe la energía derivativa en la orientación prominente.
- $\frac{\lambda_1}{\lambda_1 + \lambda_2}$ describe la circularidad local que es utilizada como un descriptor de simetría.
- $\frac{\lambda_2}{\lambda_1}$ es un descriptor de forma.

4 Detección RGB Directa.

Se implementaron todos los descriptores mencionados anteriormente, como se muestra en el siguiente listado, y se procedió a buscar puntos de máxima y mínima energía en una vecindad de 3x3x3 pixeles, la energía siempre es positiva, pero el pixel central en la máscara puede tener un valor mayor o menor que todos sus vecinos en ese intervalo y en los intervalos superior e inferior, son puntos que se mantienen constantes después del filtrado Gaussiano. Una diferencia con respecto al procedimiento sugerido por DiZenzo, es que se debe hacer un filtrado Gaussiano después de la multiplicación de las funciones derivativas expresadas en el Tensor. El procedimiento no se realizó de esta forma, como en todos los experimentos anteriores, la imagen es suavizada con el filtro Gaussiano y luego se toman sus derivadas, después se obtienen los productos y por último se obtiene el valor característico del Tensor.

5 Gradiente de Color.

La otra sección del método original de SIFT que se ha modificado es la correspondiente a la obtención del gradiente y su orientación el cual se realiza en escala de grises. Aquí solamente se procesa el plano de intensidades.

Se pueden utilizar dos métodos para obtener el gradiente y su orientación en imágenes a color. El método aproximado, donde se obtiene el gradiente total como la suma de los gradientes de cada plano RGB y el método desarrollado por Di Zenzo, ambos métodos fueron probados dando resultados muy parecidos. A continuación se muestra la función resultante que se utiliza para imágenes a color. La imagen a color se ha descompuesto en sus planos correspondientes *r_plane*, *g_plane* y *b_plane*, en la función que hace referencia a esta para el cálculo del histograma.

Las ecuaciones de Di Zenzo para la magnitud (*mag*) y orientación (*ori*) del gradiente de una imagen a son la ecuación 10 y 11:

$$ori(x, y) = \frac{1}{2} \arctan \left[\frac{2g_{xy}}{(g_{xx} - g_{yy})} \right] \quad (10)$$

$$mag(x, y) = \frac{1}{2} [(g_{xx} + g_{yy}) + (g_{xx} - g_{yy}) \cos 2(ori) + 2 g_{xy} \sin 2(ori)]^{\frac{1}{2}} \quad (11)$$

donde:

$$g_{xx} = \left| \frac{\partial R}{\partial x} \right|^2 + \left| \frac{\partial G}{\partial x} \right|^2 + \left| \frac{\partial B}{\partial x} \right|^2 \quad (12)$$

$$g_{yy} = \left| \frac{\partial R}{\partial y} \right|^2 + \left| \frac{\partial G}{\partial y} \right|^2 + \left| \frac{\partial B}{\partial y} \right|^2 \quad (13)$$

$$g_{xy} = \frac{\partial R}{\partial x} \frac{\partial R}{\partial y} + \frac{\partial G}{\partial x} \frac{\partial G}{\partial y} + \frac{\partial B}{\partial x} \frac{\partial B}{\partial y} \quad (14)$$

Estas derivadas parciales se pueden aproximar por diferencias de pixeles.

6 Resultados.

En general los resultados han sido buenos, se han logrado detectar una gran cantidad de puntos característicos en imágenes con poca iluminación, como se observa en la figura 5 y luego se ha obtenido el match entre algunos de estos puntos, dando como resultado el reconocimiento del objeto. Es importante resaltar que el mejor resultado se ha obtenido cuando se han buscado puntos de mínima energía en la vecindad 3x3x3. Además, la mayor cantidad de match points se han obtenido con el descriptor $\lambda_1 - \lambda_2$, como se muestra en la figura 6, aunque también se observan buenos resultados con el descriptor $\lambda_1 + \lambda_2$ y con el descriptor de forma λ_2/λ_1 cuyo resultado mostramos en la figura 6.

Cuando se buscaron puntos de máxima energía se detectaron pocas características en cada imagen, dando como resultado pocos match points, por otro lado cuando se buscaron puntos de mínima energía se detectaron 954 keypoints y 912 keypoints con 49 match points. Estos resultados se han obtenido en imágenes con superficie lisa. Se realizó el mismo experimento en una imagen con superficie roma y los resultados no fueron buenos, se obtuvieron una gran cantidad de keypoints pero también una gran cantidad de falsos positivos y fuera del objeto. En la figura 7 se observa el resultado después de eliminar falsos positivos eliminando keypoints que no se encuentren dentro del valor promedio del hue del canal oponente.

Se probaron varias imágenes para determinar la veracidad del descriptor, como se muestra en las figura 8 y 9, donde se comparan los resultados obtenidos con el Sift original y los resultados obtenidos con el descriptor de color $\lambda_1 - \lambda_2$.



Figura 5: Resultados obtenidos con el descriptor del Tensor de Color $\lambda_1 - \lambda_2$. En la imagen izquierda se han detectado 954 keypoints y en la imagen derecha se han detectado 912 keypoints.

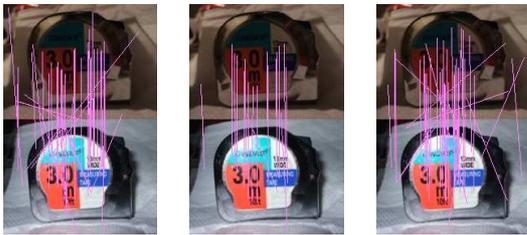


Figura 6: En la imagen izquierda se muestra el match utilizando el descriptor $\lambda_1 - \lambda_2$, con 49 match points, en la imagen central se ha utilizado el descriptor $\lambda_1 + \lambda_2$ con 21 match points y se observan menos falsos positivos, y en la imagen derecha se ha utilizado el descriptor $\frac{\lambda_2}{\lambda_1}$, con 27 match y 7 falsos positivos la detección se restringió a valores mayores a 0.5.

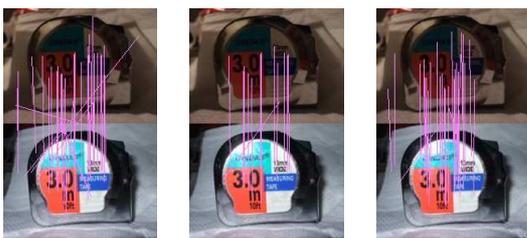


Figura 7: En la imagen izquierda se muestra el match utilizando el descriptor $\lambda_1 - \lambda_2$, y la eliminación de falsos positivos utilizando un error promedio de 0.005 para el hue, con 40 match points, en la imagen central se ha utilizado el descriptor $\lambda_1 - \lambda_2$ con 15 match points y en la imagen derecha se ha utilizado el descriptor $\frac{\lambda_2}{\lambda_1}$, con 33 match points. La detección se restringió a valores mayores a 0.1.

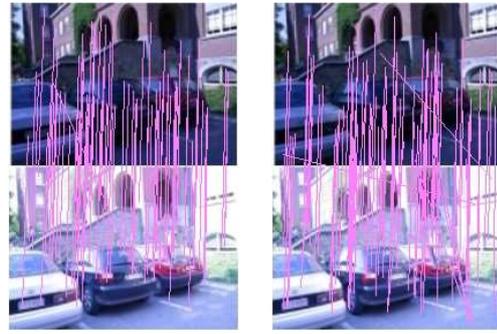


Figura 8: En la imagen izquierda se muestra el match utilizando el Sift original con 79 match points, y en la imagen derecha se ha utilizado el descriptor $\lambda_1 - \lambda_2$ con 92 match points. El descriptor de color muestra el mismo grado de robustes que el Sift original cuando se tiene buena iluminación.

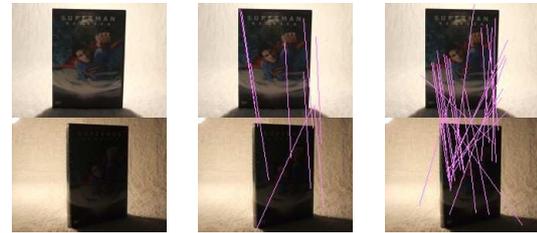


Figura 9: En la imagen izquierda se muestra el match utilizando el Sift original con cero match points, en la imagen central se ha utilizado el descriptor $\lambda_1 - \lambda_2$ con 9 match points, y en la imagen derecha se ha utilizado el descriptor $\frac{\lambda_2}{\lambda_1}$ con 26 match points, el descriptor se restringió a valores mayores de 0.5. El descriptor de color muestra ser más robusto que el Sift original cuando se tiene poca iluminación.

En general podemos determinar que si la imagen tiene buena iluminación el Sift original proporciona buenos o mejores resultados que cualquier otro descriptor, pero si la imagen tiene una pobre iluminación, el descriptor de color es un buen complemento para hacer match points entre keypoints detectados.

Referencias

- [1] Abdel-Hakim, A., Farag, A., (2006) CSIFT: A Sift Descriptor with Color Invariant Characteristics, Computer Vision and Pattern Recognition, Louisville USA.
- [2] Bosch, A., (2007) Image Clasification for Large Number of Object Categories, Univeritat de Girona.

- [3] Burghouts, G., (2007) Quality of Variant and Invariant Features, University of Amsterdam.
- [4] Di Zenzo, S., (1986) A note on the Gradient of a Multi-Image, Computer Vision Graphics, and Image Processing 33, p116-125, Italy.
- [5] Lowe, D., (2004) Distintive Image Features From Scale Invariant Keypoints, University Of British Columbia, Canada.
- [6] Ritner, L., Flores, F., Lotufo, R., (2007) New Tensorial Representation of Color Images: Tensorial Morphological Gradient Applied to Color Image Segmentation, University of Campinas, Brasil.
- [7] Schugerl, P., Sorsching R., Bailer, W., Thallinger, G., (2007) Obget Re-detection Using SIFT and MPEG-7 Color Descriptors, Institute of Information System and Information Management, Austria.
- [8] Van de Weijer, J., (2004) Color Features and Local Structure in Images, University of Amsterdam.
- [9] Van de Weijer, J., Gevers, T., (2004) Tensor Based Feature Detection for Color Images, University of Amsterdam.
- [10] Van de Weijer, J., Schmid, C., (2006) Coloring Local Feature Extraction, Gravis-Inria, Francia.