

Algoritmos de Seguimiento (“Tracking”) para aplicaciones En Realidad Aumentada.

Diego Aracena-Pizarro y Clesio Luis Tozzi
Facultad de Ingeniería Eléctrica y Computación
Departamento de Ingeniería en Computación y Automatización Industrial
Universidad Estatal de Campinas
UNICAMP - Brasil
(daracena, clesio)@dca.fee.unicamp.br

Resumen

El área de Realidad Aumentada consiste de la integración de los campos de computación gráfica, interfaz a usuario y visión computacional, para poder permitir al usuario interactuar con el mundo real a través de un ambiente real aumentado, ampliando la percepción del usuario mediante objetos virtuales insertados en escenas reales, que el usuario no es capaz de visualizar o no posee ese conocimiento. Para lograr este efecto la fase de visión computacional es de vital importancia para el funcionamiento del resto del sistema, de su éxito o fracaso, se refleja en la vista final, la que no será transparente para el usuario.

Este trabajo esta centrado en el área de visión computacional para un modelo de Realidad Aumentada, mediante la integración de herramientas que realicen en forma automática la detección de puntos característicos o corner, la determinación de la correspondencia de estos puntos en dos imágenes, el cálculo de los parámetros de movimientos y seguimiento (tracking) de los rasgos detectados, que permitan en un trabajo siguiente posicionar objetos virtuales en imágenes reales (ambientes internos y externos), sin marcas de control en el ambiente.

Utilizando la herramienta de software desarrollada, se presentan resultados con pruebas realizadas con las secuencias de imágenes, ejecutando el seguimiento de los puntos de interés a través de vistas sucesivas, calculando los parámetros necesarios para el posicionamiento de los objetos virtuales en las imágenes reales. Comparando los proceso de seguimiento con o sin uso de filtro de predicción de dimensión variable (VSDF).

Palabra claves: Realidad Aumentada, visión computacional, detección, matching, estimación de movimientos, seguimiento (tracking).

1.0 Introducción

Hace pocos años atrás, las técnicas de computación gráficas ofrecen soluciones viables para combinar imágenes reales (en video) e imágenes generadas por computador, cada vez son mas utilizadas por las ventajas que ofrecen, contar con una vista inicial que ayude en la decisión final. La composición de los objetos generados por el computador y los métodos necesarios para integrar estos campos normalmente consisten de los siguientes pasos:

- Los datos para las imágenes generadas por el computador (objetos virtuales) son preparados, mientras que el ambiente donde los objetos virtuales serán colocados deben ser seguidos (“tracking”) y/o detectados (seguimiento de características de interés).
- Posteriormente, las imágenes virtuales son transformadas del sistema de coordenadas del sistema gráfico al sistema de coordenadas de la cámara de video y combinado con las imágenes de video. Esto significa que cada imagen generada por el computador es sobrepuesta en la escena real conservando la coherencia geométrica (oclusión de objetos), como también la coherencia fotométrica (luminosidad y sombreado) entre las imágenes virtual y real.

Un modelo típico de realidad aumentada (figura 1.1) consiste de tres etapas bien diferenciadas: extracción de información usando herramientas de visión computacional, las que varían dependiendo de las necesidades de la aplicación y requerimientos del modelo; renderización basada en los atributos de la cámara, normalmente denominado generación de imágenes sintéticas; y composición combinada, la cual genera una imagen de la escena real y los objetos virtuales generados por el computador, para producir un mejoramiento de la realidad con resultados satisfactorios.

En este trabajo se centra la atención en la etapa de visión computacional, específicamente en el cálculo de los parámetros de calibración de cámara (inicialmente los parámetros externos o de movimiento) y el seguimiento (“tracking”) de los puntos o características de interés, en forma automática, que permitan en un trabajo siguiente posicionar objetos virtuales en el ambiente real, sin marcas de control (fiducial) en el ambiente. Existen pocos trabajos que apuntan en este sentido en la comunidad de Realidad Aumentada. De hecho, muchos sistemas de Realidad Aumentada desarrollados

utilizan sensores especiales y muy específicos (inerciales combinados con técnicas de visión, Azuma et al. [1999], Neumam et al. [1998]), para recuperar puntos marcados, o calcular de manera mas o menos interactiva la posición de estos con la ayuda del usuario (Vallino et al. [1998], Berger et al. [1998-1999]). Otras técnicas requieren colocar en el ambiente puntos de control pasivos (fiduciales) para señalar el ambiente (círculos, cuadrados o rectángulos) que sobresalgan en la imagen (Koller et al. [1998], Neumam et al. [1996-1998], Brantner [1998], Seo et al. [2000]), pero intervenir el ambiente no siempre es posible. Las fases de generación de imágenes sintéticas, renderización y composición combinada no son descritas y pueden ser revisadas en Berger et al. [1995], Collins et al. [2000], Koller et al. [1998], Tuceryan et al. [1995].

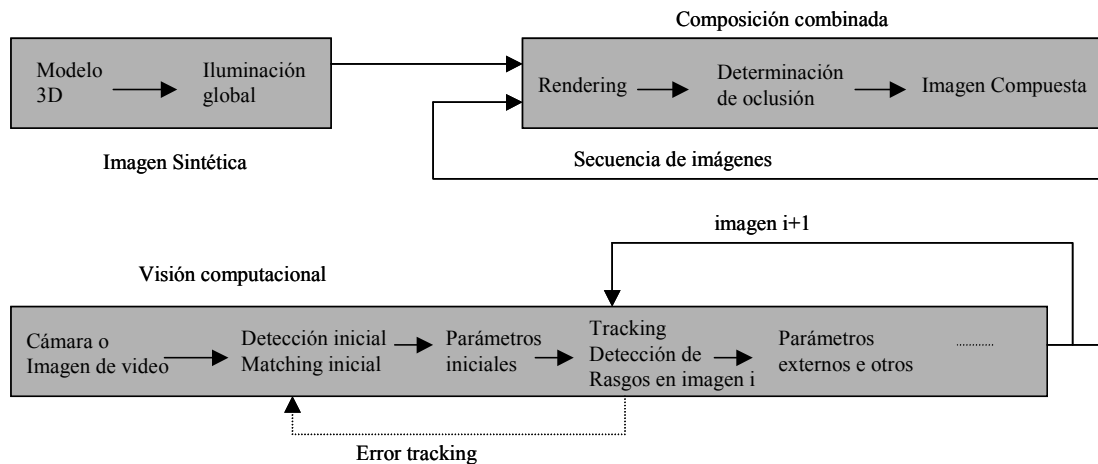


Figura 1.1 Arquitectura típica de un sistema de Realidad Aumentada.

Como consecuencia del uso de las herramientas desarrolladas, se presentan resultados con las secuencias de imágenes, realizando el seguimiento de los puntos de interés a través de vistas sucesivas, calculando los parámetros necesarios para el posicionamiento de los objetos virtuales en las imágenes reales, comparando los procesos de seguimiento con o sin el uso de filtro de predicción de dimensión variable (VSDF) (McLauchlan et al. [1995]).

2.0 Proceso de Calibración

El proceso de calibración de cámara consiste en determinar la geometría y características internas de la cámara (parámetros intrínsecos, tamaño horizontal ($1/s_u$) y vertical ($1/s_v$) de los píxeles, coordenadas de proyección $O'(r_0, c_0)$ del centro óptico), la posición y orientación de la cámara relativa a una imagen padrón con un sistema de coordenadas (parámetros extrínsecos). Estos parámetros normalmente son calculados desde un padrón de calibración que contiene rasgos fácilmente detectables de manera precisa en la imagen capturada (Tuceryan et al. [1995]).

El proceso de calibración para la determinación de los parámetros intrínsecos y extrínsecos, no necesita ser utilizado para el cálculo de la posición o punto de observación de la cámara en forma repetitiva, porque los parámetros intrínsecos permanecen prácticamente los mismos en toda la secuencia, debido a que la distancia focal no varía. Los parámetros extrínsecos varían cada vez que se produce un desplazamiento de la cámara y pueden ser determinados por métodos de estimación de movimiento (sección 3). Los parámetros intrínsecos se deben determinar al comienzo de la secuencia de trabajo, utilizando una grilla padrón.

Considerando las ecuaciones de perspectivas de un sistema de referencia padrón de cámara, asumiendo que (r_0, c_0) son el centro del sistema de coordenadas del plano imagen O' , s_u y s_v reflejan la naturaleza no isotrópica de las cámaras (ejemplo, cámaras con píxeles no regulares) y el aspecto proporcional de la cámara. Sean $f_u = s_u f$ y $f_v = s_v f$, reduciendo los parámetros intrínsecos de la cámara a f_u, f_v, r_0, c_0 y los parámetros extrínsecos o del movimiento a R y T . Logrando relacionar un punto 3D $P=(x,y,z)$ y su proyección (u, v) de la forma

$$\begin{pmatrix} s \cdot u \\ s \cdot v \\ s \end{pmatrix} = \begin{pmatrix} f_u & 0 & c_0 & 0 \\ 0 & -f_v & r_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

El proceso de calibración requiere conocer puntos 3D y su proyección correspondiente en puntos de imagen 2D, estableciéndose un sistema de ecuaciones que permiten determinar los parámetros intrínsecos y extrínsecos por el empleo del método de minimización por cuadrados mínimos. Para mas detalles de este proceso, ver Faugeras [1993], Shapiro et al. [2000], Trucco et al. [1998], Turcercyan et al. [1995] o Aracena y Tozzi [2001].

3.0 Estimación de movimientos

El problema de estimar movimientos tuvo su mayor atención a partir de la introducción de la aproximación de la matriz esencial propuesta por Longuet-Higgins et al. [1981]. Donde se demuestra que el desplazamiento 3D relativo de una cámara u objeto en movimiento se puede recuperar a través de las restricciones geométricas entre dos imágenes de la misma escena. De esta aproximación resultan algoritmos con características bastantes favorables: la heurística no precisa contar con un conocimiento a priori de la escena; son computacionalmente sencillos (comparados con los algoritmos de estimación de movimiento no intrínsecos); y estan mas centrados en el uso de técnicas basadas en álgebra lineal (ver, Tsai et al. [1984], Faugeras [1993], Fusiello et al. [1998], Zhang et al. [1996], Hartley [1995-1997]).

De manera resumida la ecuación de Longuet-Higgins et al. [1981], se aplica cuando se utilizan coordenadas normalizadas y cámaras calibradas. Si el movimiento entre dos posiciones de la cámara es descrito por la rotación R y la translación T , determinados a partir de puntos correspondientes (m_1 puntos de la primera imagen y m_2 puntos de la segunda imagen), luego las restricciones coplanares relativas a las líneas epipolares con los puntos correspondientes en ambas imágenes son escritas como:

$$m_2 \cdot (T \times Rm_1) = m_2^T E m_1 = 0$$

La matriz E (esencial) es el producto de la matriz ortogonal R y la matriz anti-simétrica T_{\wedge} , con determinante $Det(T_{\wedge}) = 0$ y $rank = 2$ definida hasta un factor de escala (Zhang et al. [1996], Trucco et al. [1998], Fusiello et al. [1998]).

La matriz esencial y la matriz fundamental están relacionadas por la matriz de calibración de la forma $F = K^{-1T} E K^{-1}$, donde K es la matriz de calibración con los parámetros intrínsecos de la cámara y F la matriz fundamental con $rank = 2$, $norm(F) = 1$ y definida también hasta un factor de escala.

La técnica utilizada en este trabajo para determinar la matriz fundamental y posteriormente la matriz esencial, consiste en considerar un conjunto suficientemente de puntos correspondientes (mas de lo necesario)

$\{(m_{1i}, m_{2i}) | i = 1, \dots, n\}$ en las imágenes. Ellos deben satisfacer la ecuación epipolar $\tilde{m}_{2i}^T F \tilde{m}_{1i} = 0$ que se puede describir por un sistema de ecuaciones de la forma $u_i^T f = 0$, ver Zhang et al. [1996], Faugeras [1993] y Trucco et al. [1998].

La solución, utilizando autosistemas, es la columna de la matriz de autovectores correspondiente al menor autovalor positivo distinto de cero. Normalmente este método es denominado algoritmos de los 8 puntos correspondientes (Hartley [1997], Zhang et al. [1994], Trucco et al. [1998]).

Posteriormente forzar la restricción de $rank = 2$, utilizando la descomposición en valores singulares de la matriz fundamental. (ver Ma et al [1999], Hartley [1997], Zhang et al [1996]).

El criterio lineal de Zhang et al. [1996], que consiste en minimizar una función distancia que minimiza la matriz fundamental, permite mejorar el resultado de la matriz fundamental e indirectamente lograr parámetros de movimientos mas precisos.

Utilizando la propiedad que $\tilde{m}_2^T F \tilde{m}_1 = \tilde{m}_1^T F^T \tilde{m}_2$, se puede establecer una función criterio basadas en las distancias euclidianas de los puntos a sus respectivas líneas epipolares como:

$$\min_F \sum_i w_i^2 \left(\tilde{m}_{2i}^T F \tilde{m}_{1i} \right)^2,$$

Hartley [1995-1997] observa que la ecuación de Longuet-Higgins para la matriz fundamental esta mal condicionada, porque el valor de los píxeles representado en forma homogénea son significativamente superiores a 1, por lo que propone una normalización isotrópica que consiste en trasladar todos los puntos, tal que su centroide sea el origen y luego sean escalados tal que la distancia promedio desde el origen sea $\sqrt{2}$. Posteriormente determinar la matriz fundamental bajo esta normalización de los puntos imagen.

Un problema vital en la estimación de movimiento y seguimiento es el proceso de determinar la correspondencia entre los puntos identificados entre dos imágenes en la secuencia. Este problema es conocido como matching de puntos característicos, puntos de interés o "tokens". Los cuales pueden ser determinados en forma manual o automática.

4.0 Determinación de rasgos y correspondencia entre dos imágenes

La etapa de detección de puntos característicos se basa en la utilización del algoritmo de Harris y Stephens [1988]. Ellos emplean la función de auto correlación de la imagen para calcular una medida que detecta la presencia de un punto de interés. El método consiste en calcular las gradientes espaciales de la imagen $[f_r, f_c]^T$, posteriormente se considera una vecindad W alrededor de un punto p , formando una matriz \hat{C} ponderada del momento de segundo orden definida en Trucco et al. [1998]. Esta matriz caracteriza la estructura de los niveles de gris a través del cálculo de los autovalores de la matriz \hat{C} y su interpretación geométrica. Considerando λ_1 y λ_2 como los autovalores máximos y mínimo de la matriz \hat{C} , respectivamente, se puede clasificar la estructura de la imagen entorno al píxel por estos autovalores. Se construye un histograma con los valores de λ_2 s determinándose un threshold entre un 10 a un 20% del valor máximo, que será utilizado como discriminador en la selección de los puntos de interés con $\lambda_2 > \tau$. Una vez detectados los puntos de interés en dos imágenes se puede determinar la correspondencia entre los puntos detectados en ambas imágenes, a través de un algoritmo de matching.

Como algoritmo de matching, se utiliza la propuesta realizada en Pilu [1997] que usa como medida local un valor de confianza determinado por la correlación normalizada C_{ij} entre cada uno de los puntos de interés de ambas imágenes. Para su cálculo se considera las intensidades en una área centrada en I_i y J_i , siendo los rasgos de la primera y segunda imagen, respectivamente. Este resultado es integrado a un algoritmo de proximidad relatado en Scott y Longuet-Higgins [1991], estableciéndose como función objetivo para el cálculo de la matriz de similitud, la siguiente ecuación

$$G_{ij} = \frac{(C_{ij} + 1)}{2} e^{-r_{ij}^2 / 2\sigma^2} \quad i = 1 \dots m, j = 1 \dots n$$

donde cada elemento G_{ij} es la distancia con peso gaussiano entre los rasgos I_i y J_i correspondientes a la primera y segunda imagen, respectivamente. $r_{ij} = \|I_i - J_j\|$ es la distancia euclidiana. G es positiva y decrece monotónicamente entre 1 y 0 con la distancia. El parámetro σ controla el grado de interacción entre los dos conjuntos de rasgos.

Este algoritmo es bastante eficiente, simple y con buen nivel de desempeño, que lo hace apropiado para aplicaciones de Realidad Aumentada. El proceso de matching en el modelo de Realidad Aumentada tiene dos variantes; en la fase inicial actúa en toda la imagen y en el ciclo de seguimiento colabora en la detección del matching inicial por correlación corroborando la detección por las medidas de similitud de los puntos característicos en la ventana de investigación. El número puntos casados y la estrategia de continuidad del ciclo de seguimiento, permitirán determinar si el ciclo continua o debe regresar a la fase inicial de detección y matching.

5.0 Seguimiento (“tracking”)

Los esquemas de seguimientos que muestra la figura 5.1, configurados para evaluar pruebas de la fase de visión computacional, considerando movimiento de una cámara y escena estática, las variantes del ciclo de seguimiento consisten en la estimación de la posición de la cámara por dos vistas (método de normalización isotrópica de Hartley [1997], complementado con la solución de la matriz de fundamental de Zhang et al. [1996], extrayendo los parámetros de movimiento por el método de factorización de Zhuang et al. [1986]) y la estimación mediante un filtro recursivo de predicción.

Para el esquema de seguimiento con filtro recursivo, se utiliza en esta aplicación un filtro de dimensión variable de los estados (VSDF) (Bar-Shalom [1993], Welch et al. [2001], Brown [1997], McLauchlan[1994]), que consiste en determinar un conjunto de estados globales como un vector x , y un conjunto de estados locales a ser estimados $y_i, i = 1 \dots n$. Todo el conjunto de parámetros globales y locales forman el vector total de estado y están relacionados por la observaciones $z_i(j)$ en el tiempo j por la ecuación de medida

$$z_i(j) = h_i(j; x; y_i) + w_i(j)$$

donde $w_i(j)$ es un vector de distribución Gaussiana con media cero y covariancia $R(j)$. Como cualquier algoritmo de filtro de Kalman, las reglas de actualización son derivadas de

$$P^{-1} \leftarrow P^{-1} + H^T R^{-1} H, \quad \hat{x} \leftarrow \hat{x} + x d, \quad (5)$$

donde P es la matriz de covariancia (o de información) y $xd = PH^T R^{-1}v$. Aquí H es el jacobiano de la matriz $\frac{\partial h}{\partial x}$.

Ambos H y h son evaluados en el valor anterior de \hat{x} . La actualización de P es realizada calculando antes. En (5) la expresión xd no es mas que la ganancia Kalman en función de la matriz de covariancia. Mayores detalles con respecto al filtro VSDF en McLauchlan [1994-1999], Nickel [2000], Brown [1997].

Para el filtro recursivo se considera las siguientes definiciones para su implementación: los estados globales que representan los parámetros de movimiento $x = (\theta, \omega, \kappa, T_x, T_y, T_z)$, donde los ángulos θ, ω, κ son los ángulos de Euler que representan la matriz de rotación R , y (T_x, T_y, T_z) representan la translación en los ejes x, y, z ; los estados locales corresponden a los puntos 3D $P = (X, Y, Z)$; $y = P$, que es la proyección de la medida en el plano imagen; y la ecuación

de medida corresponde a la ecuación de perspectiva $z = (u_i, v_i)^T$, definida como: $h(u, v) = \left(\frac{R_1 P_1 + T_x}{R_3 P_1 + T_z}, \frac{R_2 P_1 + T_y}{R_3 P_1 + T_z} \right)^T$

donde las R_i son las filas de la matriz de rotación.

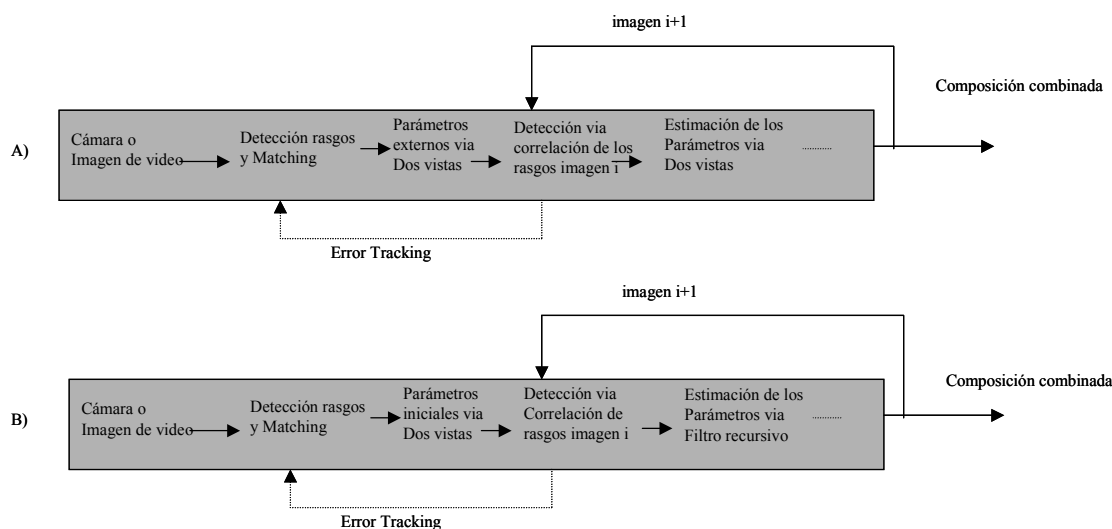


Figura 5.1. Modelo de la fase de visión computacional a) estimando la posición futura mediante dos vistas b) Estimando la posición futura con filtro recursivo.

Para implementar la actualización recursiva del filtro VSDF se necesita calcular los jacobianos de h con respecto a x y con respecto a y , los cuales fácilmente son obtenidas por la diferenciación, detalles en McLauchlan [1995-1999].

6.0 Resultados y discusión

Hasta el momento se ha implementado los diferentes algoritmos descritos en la fase de visión computacional como muestra la figura 5.1a y 5.1b. a nivel de prototipo, utilizando el lenguaje de programación MatLab (<http://www.mathworks.com>) realizando pruebas con los diferentes módulos. Las imágenes sintéticas fueron generadas en PovRay y las imágenes reales adquiridas en ambientes internos y externos con una cámara digital de bajo costo Microtek (tipo WebCam).

La efectividad de cualquier herramienta que involucre seguimiento (“tracking”) se mide directamente por la capacidad que posee esta en seguir los rasgos en la secuencia de imágenes. Pero este seguimiento es efectivo cuando las etapas previas, las que permiten su inicialización, están bien ajustadas e introducen un mínimo de error a la tarea de seguir.

La estrategia adoptada en la fase inicial de detección y matching, consiste en presentar al usuario los rasgos o puntos de interés detectados y casados automáticamente, para su elección de acuerdo a la conveniencia de la aplicación. La característica automática de esta fase cobra su importancia, cuando aparecen nuevos rasgos y el usuario pueda interactuar para anexar mas al proceso o descartar algunos conflictivos.

En la fase de detección de rasgos fue empleado el modelo propuesto por Harris y Stephens [1988], con modificaciones para una mejor operación. Las modificaciones consistieron en ampliar la posibilidad de uso con diferentes mascarar gradientes Harris, Robert, Sobel, y Gaussiana (Schmid et al. [1997]) y la medida de confianza para la selección

de los rasgos. En las pruebas fueron utilizadas las mascararas gradiente de Sobel y Robert, por presentar una mejor precisión en la detección de corner y puntos de interés con las imágenes escogidas. También, se debe considerar que el tamaño de la ventana padrón o template, incide significativamente en la efectividad del detector, influyendo en la magnitud de la medida de confianza y por consiguiente en la determinación del valor threshold de selección. En el caso en que la medida de confianza (menor autovalor de la matriz ponderada (Shi et al. [1994], Trucco et al. [1998]) es mayor que el threshold, el punto es considerado como distinto y agregado a la lista de puntos de interés. La lista es refinada eliminando los puntos vecinos por supresión no máxima. En las pruebas se utilizó una ventana template de 7×7 para el cálculo de la medida de confianza y otra de 13×13 para la eliminación de los puntos vecinos.

Para determinar la correspondencia entre los puntos detectados en dos imágenes de la secuencia, fue empleado el método de matching presentado por Pilu [1997], quien realiza una optimización del algoritmo de matching presentado por Scott y Longuet-Higgins [1991]). Este algoritmo emplea como medida de confianza local la correlación normalizada (NCC) y como medida de proximidad la distancia euclidiana entre los rasgos, construyendo una matriz de similaridad que sirve para indicar la correspondencia de los rasgos entre las imágenes. Esta correspondencia se establece cuando se logra un máximo en filas y en columnas (dos maneras) (Sanchiz et al. [1999]). La figura 6.1 muestra una parte de las imágenes casadas con sus puntos correspondientes. Nótese que existen errores en el matching de algunos puntos, influenciados principalmente por la luminosidad que incide en la foto, permaneciendo estable en los puntos que mantienen las condiciones de intensidades previas captadas en la imagen correspondiente, sin variar sus parámetros ante el movimiento, el mismo problema se aprecia en imágenes tomadas con imágenes reales en ambientes externos.

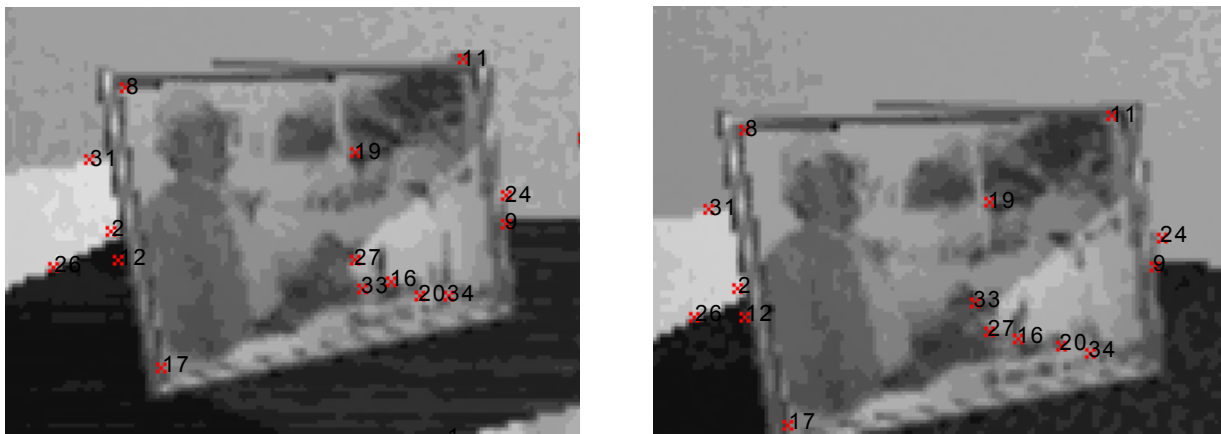


Figura 6.1 Matching entre dos imágenes tomadas en diferentes posiciones.

En todas las pruebas realizadas en la fase de detección y matching, el algoritmo de matching presenta una tasa de error menor que el 20%, considerándose un método apropiado para ser utilizado en seguimiento. Este porcentaje no es significativo cuando es utilizado como entrada a un método de dos vistas robusto (utilizando métodos estadísticos) para el cálculo de los parámetros de movimiento (Schmidt et al. [2001]). Por otro lado, aun no existen propuestas que identifiquen los puntos errados y su eliminación de manera automática (Berger et al. [1999]). Por lo tanto, es claro que la estrategia de dejar al usuario escoger los puntos de interés propuestos por el sistema, permite mediante observación visual, eliminar los puntos errados en esta etapa.

A partir de los puntos correspondientes disponibilizados en la etapa detección y matching se puede calcular la matriz fundamental, esencial y finalmente llegar a determinar los parámetros de movimiento o calibración extrínsecos. Para el cálculo de la matriz fundamental se utiliza el método de translación y escala isotrópica descrito en Hartley [1997], optimizando la matriz fundamental a través del método no lineal que la minimiza de acuerdo a la distancia entre los puntos y las líneas epipolares (Zhang et al. [1996]). Una vez determinada la matriz fundamental se puede calcular la matriz esencial para extraer los parámetros del movimiento (matriz de rotación y el vector de translación). En la fase de extracción de los parámetros de movimiento (factorización) se utiliza el método propuesto por Zhuang et al. [1986]. Cabe hacer notar que los métodos de factorización presentado en Haralick et al. [1993] y Faugeras [1993] presentan resultado muy similares, lo que torna mas relevante la determinación de la matriz fundamental. Los parámetros de la matriz que representan el movimiento son importantes para determinar la posición y orientación de la cámara, por reconstrucción de perspectiva establecer donde

será intervenida la imagen real y como estimación inicial para los parámetros que representan el movimiento en el filtro de predicción.

En las pruebas de operación realizadas siempre fue considerado estimar la posición futura de los rasgos a partir de la información actual, este método difiere de los seguimientos de rasgos que comienzan con la primera imagen y posteriormente realizan la búsqueda para cada rasgo en la segunda imagen, utilizando una ventana de búsqueda centrada en la posición de un punto característico de la primera imagen, asumiendo que el movimiento si tubo lugar, no supera el tamaño de ventana de investigación. Este mecanismo normalmente es utilizado cuando se trabaja con flujo óptico por correlación (Vasquez [1996], Neumam et al. [1998]). También, difiere del modelo de seguimiento por matching sucesivo, es decir, para cada imagen nueva en la secuencia se detectan los puntos de interés y se realiza el proceso de matching entre las dos ultimas imágenes (Berger et al. [1999]). El abordaje aquí presentado considera un movimiento aleatorio desconocido en que se estima la posición futura de los rasgos con el conocimiento de los parámetros de movimiento actual. En ambos casos se debe considerar la reconstrucción de proyección o reconstrucción 3D, que consiste en determinar por minimización la posición de los puntos imagen en coordenadas de cámara en el sistema de coordenada del mundo.

Para el cálculo de la reconstrucción de proyección o reconstrucción 3D, existen formulaciones sencillas que consideran restricciones con respecto al comportamiento de la rotación y la característica del desplazamiento de la translación, los que deben ser significativos, porque antes movimientos pequeños o próximos a una superficie degenerada (ejemplo, plana), desafortunadamente los resultados son muy insatisfactorios (Zhang et al. [1998]).

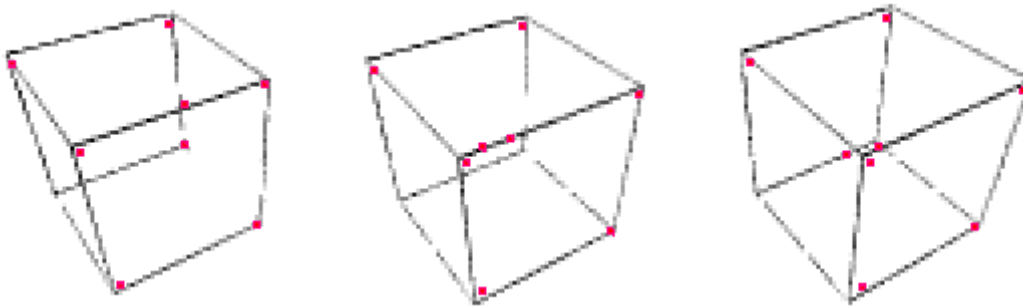


Figura 6.2 Secuencia del algoritmo de tracking sin filtro de las imágenes 1ª, 5ª y 14ª.

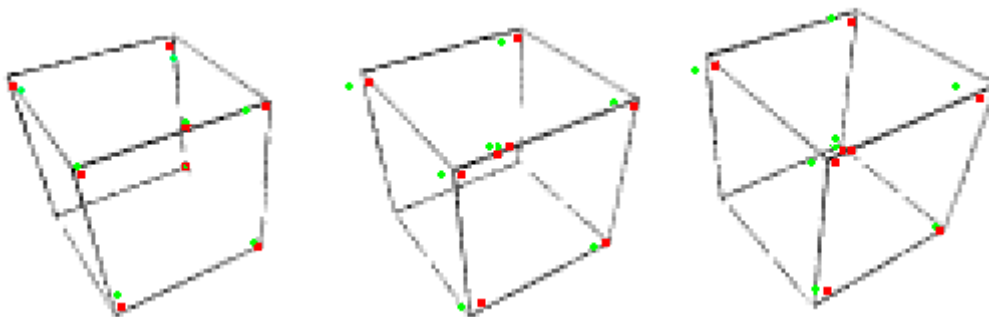


Figura 6.3 Secuencia del algoritmo de tracking con filtro recursivo de las imágenes 1ª, 5ª y 14ª.

Es por esta razón que el seguimiento sin considerar filtro recursivo tiene un comportamiento inesperado, cuando el movimiento en general es pequeño, en el caso que el movimiento sea significativo la estimación de la posición de los rasgos pronosticados tienen un error pequeño del orden 1 a 2 píxeles en su mayoría o coinciden exactamente en algunos casos. Este comportamiento inesperado es atribuible a la influencia del factor de escala introducido en el cálculo de los parámetros del movimiento. El proceso de verificación, matching por correlación normalizada, permite corregir los errores introducidos por el factor de escala, dejando los rasgos mas determinados y listos para ser utilizados en la etapa siguiente.

Para el caso de seguimiento con filtro, si el movimiento no tiene un cambio abrupto es capaz de seguir los rasgos en forma satisfactoria, su convergencia esta mas ligada a la calidad del proceso de matching por correlación normalizada que actúa como una realimentación en el proceso de seguimiento. Las pruebas de operación se presentan en las figuras 6.2 y 6.3, para ambos casos descritos.

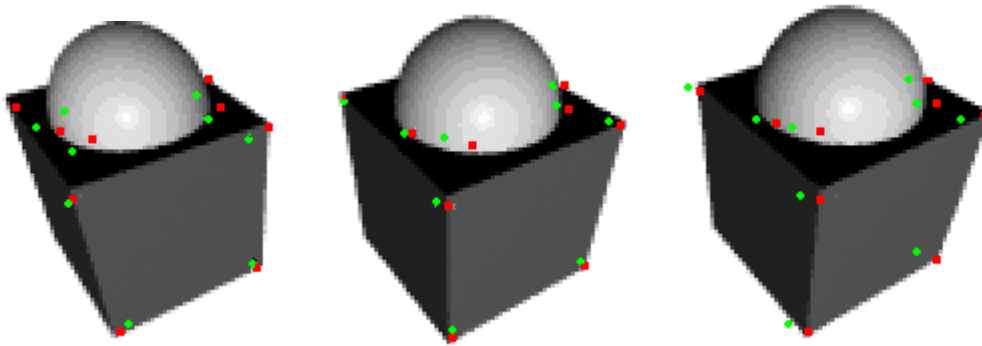


Figura 6.4 Secuencia del algoritmo con filtro recursivo en el seguimiento de 12 imágenes.

Se puede notar que el seguimiento o tracking usando filtro es capaz de seguir la secuencia de imágenes, salvo si se presenta un cambio brusco del punto de observación en la secuencia. La figura 6.4 presenta la secuencia con una imagen compuesta por puntos que varían de intensidad dependiendo de la luminosidad. La aplicación del algoritmo sin considerar el filtro tiene un comportamiento errático, mientras que con la utilización del filtro recursivo, los errores introducidos por esos pocos puntos inestables no producen grandes variaciones lo que posibilita mantener el seguimiento en la secuencia.

La inestabilidad introducida por los puntos poco definidos e influenciados por los cambios de intensidad, que varios autores han reportado como punto débil del algoritmo de dos vistas, demanda que la estimación de movimiento se realice con un filtro recursivo que permita minimizar este efecto.

7.0 Conclusiones

Para la implementación de los esquemas de seguimiento que se presentan en 5.1a y 5.1b, se considero utilizar métodos rápidos y que presenten una operación apropiada para solucionar el problema de seguir rasgos, donde el movimiento es desconocido y aleatorio para el caso de Realidad Aumentada. Esto último condiciona la selección de los algoritmos en todas las fases (detección, matching y ciclo de seguimiento). Se considera el seguimiento de puntos de interés o corner, por ser rasgos que poseen las características mas relevantes tales como distinción, unicidad, robustez, invariabilidad y estabilidad, lo hacen apropiado para esta aplicación.

Del estudio de los métodos de detección de puntos de interés o corner (Ballart et al. [1983], Haralick et al. [1993], Trucco et al. [1998], Fusch et al. [1998], Guerrero et al. [1998], Shapiro et al. [2000]), se puede decir que los métodos donde están involucrados los cálculos de las derivadas de primer orden (gradientes espacial de la imagen) son mas rápidos y presentan un mejor comportamiento con imágenes reales en ambientes externos, que los métodos que calculan curvatura a través de derivadas de orden superior, siendo estas ultimas mas precisa en imágenes reales en ambientes internos. El detector de corner o puntos de interés de Harris y Stephens [1988] fue escogido por presentar buen desempeño en la detección de rasgos en imágenes reales en ambientes internos y externos.

El algoritmo de matching de Pilu[1997] fue escogido por ser un método que considera medidas locales y de exclusión por proximidad, resultando un algoritmo rápido, sencillo y efectivo en la determinación de la correspondencia de puntos bien definidos, no influenciados por los cambios de intensidad antes el movimiento de la cámara. Como también, el algoritmo difiere del matching por correlación normalizada tradicional (Jain et al. [1995], Zhang et al. [1994], Fusiello et al. [1998]), porque brinda un buen comportamiento ante la rotación y cambios de escala.

Para la estimación de movimientos, se prioriza el cálculo de la matriz fundamental por poseer mayor información, ya que encapsula la matriz esencial y por consiguiente la geometría epipolar y los parámetros extrínsecos del movimiento. El algoritmo de dos vistas de los 8 puntos correspondientes es la mejor alternativa para decodificar la matriz y utilizar su resultado para la estimación de movimientos.

El algoritmo de seguimiento con filtro recursivo que sirva para aplicaciones en realidad aumentada, se precisa contar con un filtro que ofrezca la posibilidad de cambiar los estados de la ecuación de medida, esta alternativa no es fácil de implementar con el Filtro de Kalman Extendido para este caso no lineal, dado que el cambio de estados involucra normalmente reconfigurar todas las matrices que representan de actualización del filtro, donde no es trivial las incorporaciones y eliminaciones de los estados, manteniendo un nivel operación estable y que el algoritmo pueda convergir logrando predicciones mas precisas. Mientras que el filtro recursivo VSDF de McLauchlan et al. [1995] ofrece esta operación de actualización en forma apropiada e inclusive de manera mas sencilla que trabajar con filtro de estados variables

de Schmidt-Kalman (Brown et al. [1997]). Facilitando la tarea de incorporar, eliminar estados y con una lógica de operación similar a un filtro de Kalman Extendido Implícito.

Los modelos de seguimientos descritos tienen una operación bastante apropiada a nivel de prototipo en Matlab, lo que permite inferir que son apropiados para operar en tiempo real, para resoluciones de imágenes de 320×240 . Para resoluciones mayores el tiempo crece exponencialmente.

Hasta el momento se ha trabajado en la fase de visión principalmente, porque se considera que es de vital importancia para el funcionamiento del resto del sistema, en la actualidad existe una atención especial en solucionar de la mejor manera la etapa de visión computacional, porque es la parte central de cualquier aplicación en Realidad Aumentada, de su éxito o fracaso, se refleja en la vista final, la que no será transparente para el usuario (Azuma et al. [1999]).

Como trabajo futuro, se espera seguir realizando prueba con imágenes reales en ambientes externos e ir avanzando en el afinamiento de los diferentes módulos, hasta lograr la integración de todas las fases del modelo de Realidad Aumentada descrito, anexando la fase de combinación en aplicaciones de Realidad Aumentada en ambientes externos, por ejemplo, alteración urbana, Museos, sitios arqueológicos o herencia virtual.

8.0 Referencias

- Aracena, D. Tozzi, C. [2001] "Realidad Aumentada Integrando Visión Computacional" Tutorial Universidad San Agustín. Arequipa Perú.
- Azuma, Ronald [1997] "A Survey of Augmented Reality", In Presence: Teleoperators and Virtual Environment, August, pp.355-385.
- Azuma, Ronald [1993] "Tracking Requirement for Augmented Reality", Communications of the ACM, Vol.36, N°7, July.
- Ballard, D.H. and Brown, C.M. [1983] "Computer Vision", Prentice Hall, ISBN 0-13-165316-4.
- Bar-Shalom, Y. Li, X. [1993] "Estimation and Tracking: principles, techniques and software", Artech House, Boston.
- Brantner, S. [1998] "A Vision-Based Tracking technique for Augmented Reality, Master Dissertation in Technischen Universität Graz, Germany.
- Berger, M-O, Wrobel-Dautcourt, B. Petitjean, S. And Simon, G.[1999] "Mixing synthetic and video images of an outdoor urban environment", Machine Vision and Applications, Vol (1999), (11), pp. 145-159, Nov. 1999.
- Brown, R. G. And Hwang, Y. C. [1997] "Introduction to Random Signals and Applied Kalman Filtering", 3th Edition, John Wiley & Sons, Inc. 1997.
- Fusiello, A. Censi, A. Roberto, V.[1999] "Image Stabilization by Features Tracking", in Proceeding of the 10th International Conference on Image Analysis and Processing, pp.665-667. Venice, Italy.
- Collins, R. And Dellaert, F.[1998] "Fast Image-Based Tracking by Selective Pixel Integration", in <http://fizbin.eecs.lehigh.edu/~tboult/FRAME/dellaert/framerate.html>
- Faugeras, O. [1993] "Tridimensional Computer Vision: A Geometric Viewpoint" Artificial Intelligence. MIT Press.
- Forstner, W. [1994] "A Framework for Low Level Feature Extraction", Lecture Notes in Computer Science, Vol. 802, Computer Vision – ECCV'94. Pp. 383-394.
- Fusiello, A. Tommasini, T. Trucco, E. And Roberto V. [1998] "Making Good Features to Track Better", in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Pp. 178-183, Santa Barbara, CA, 1998.
- Fuchs, C. And Heuel, S. [1998] "Feature Extraction", Report Research Institute for Photogrammetry University Bonn, February, claudia@ipb.uni-bonn.de.
- Guerrero, J.J. and Sagüés, C. [1998] "Tracking Features with Camera Maneuvering for Vision-Based Navigation", Journal of Robotic Systems, Vol. 15, (4), pp: 191-206, 1998.
- Hager, G. And Toyama, K. [1998] "X-Vision: A Portable Substrate for Real Time Vision Applications", CVIU: Computer Vision and Image Understanding, Vol.69, (1), pp: 23-37, January 1998.
- Haralick, R.M. and Shapiro, L. [1992] "Computer and Robot Vision", Ed. Addison-Wesley Publishing Company, Vol. 1, 1992.
- Haralick, Robert and Shapiro, Linda [1993] "Computer and Robot Vision", Addison Wesley, Vol 2, pag. 212-227.
- Harris, C. And Stephens, M.[1988] "A Combined Corner and Edge Detector", In Proceeding 4th Alvey Vision Conference (AVC88) pp. 147-151.
- Hartley, R. I. [1995] "In Defence of the 8-point Algorithm", in Proceeding of the IEEE International Conference on Computer Vision, 1995.
- Hartley, R. I. [1997] "Kruppa's Equations Derived from Fundamental Matrix", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, (2), pp. 133-135, Feb. 1997.
- Jain, R. Kasturi, R. And Schunck, B. [1995] "Machine Vision", McGraw-Hill and MIT Press, New York 1995.

- Koller, D. Klinker, G. Rose, E. Breen D. Whitaker R. e Tuceryan M. [1998]"real-time Vision-Based Camera Tracking for Augmented Reality Applications", Tec. Rep. California Inst. Of Technology, e-mail dieter.koller@autodesk.com
- McLauchlan, P. and Murray D. [1995] "A Unifying Framework for Structure and Motion recovery From Image Sequence", In Proc. 5th Int'l Conf. On Computer Vision, Boston, pp. 314-320, June 1995.
- McLauchlan, P. [1999] "The Variable State Dimension Filter", Technical Report VSSP 4/99, University of Surrey, Dept. of Electrical Engineering, Nov. 1999.
- Neumann, U. You, S [1999] "Natural Feature Tracking for Augmented Reality", IEEE Transactions on Multimedia, Vol.1, 1, pp. 53-64.
- Neumann, U. And Park, J. [1998] "Tracking for Augmented Reality on Wearable Computers", Virtual Reality Journal, 3, pp. 167-175. Springer Verlag, London Ltd.
- Pilu, M. [1997] "A Direct Method for Stereo Correspondence based on Singular Value Decomposition", in Proc. CVPR '97, pp. 261-266.
- Ravela, S. Draper, B. Lim, J. And Weiss, R.[1996] "Tracking Object Motion Across Aspect Changes for Augmented Reality", in Proc. ARPA Image Understanding Workshop, Palm Spring, CA, 1996.
- Sanchiz, J.M. and Pla, F. [1999] "Feature correspondence and Motion recovery in vehicle planar navigation", Pattern Recognition, Vol. 32, pp: 1961-1977, 1999.
- Sanchiz, J.M. Marchant, J.A. Pla, F. And Hague, T. [1998]"Real Time Visual Sensing for Task Planning in a Field Navigation Vehicle", Real Time Imaging, Vol. 4, pp: 55-65, 1998.
- Scott, G. And Longuet-Higgins, C. [1991] "An Algorithm for Associating the Features of Two Images", in Proc. Royal Society London, Vol B244, pp. 21-26, 1991.
- Schmidt, J. Scholz, I. And Niemann, H. [2001] "Placing Arbitrary Objects in Real Scene Using a Color Cube for Pose Estimation", in Pattern Recognition DAGM Symposium, LNCS 2191, pp. 421-428, München, Germany.
- Seo, Y. And Hong, S. [2000] "Calibration-Free Augmented Reality in Perspective", IEEE Transactions On Visualization and Computer Graphics, Vo. 6,(4), pp: 346-359, October-December 2000.
- Shapiro, Linda and Stockman, George , [2000] "Computer Vision" updated in March 2000 , <http://george.ee.washington.edu/courses.html>
- Shi, J. And Tomasi, C. [1994] "Good Features to Track", IEEE Conf. On Computer Vision and Pattern Recognition, Seattle, June 1994.
- Sun, C. And Wu, X.[1997] "A Method for Automatic Segmentation of Fiducial Mark", in International WorkShop on Image Analysis and Information Fusion, pp. 43-49, Adelaide Australia.
- Tsai, R. And Huang, T. [1994] " Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 6,(1), pp. 13-26, Jan.
- Trucco, E. And Verri, A. [1998] "Introductory techniques for 3D Computer Vision", Prentice Hall, NJ.
- Tuceryan, M. Greer, D. Whitaker, R. Breen D. Crapton C. and Ahlers K.[1995] "Calibration Requirements and Procedures for a Monitor-Based Augmented Reality System", IEEE Trans. On Visualization and Computer Graphics, Vol. 1, 3, Sep. Pag 255-273.
- Vallino James R. [1998] "Augmented Reality with Minimal Calibration", <http://www.cs.rit.edu/~jrv/research/ar/index.html>.
- Vázquez, Fernando [1996] "Segmentación de Imágenes en grafos de contorno. Aplicación a la estimación de la profundidad y el movimiento relativo para un robot móvil autónomo", Tesis Doctoral Universidade de Vigo España, Junio.
- Welch, G. And Bishop, G. [2001] "Kalman Filter", SIGGRAPH 2001, Los Angeles, CA, August 2001.
- Weng, J. Huang, T. and Ahuja, N. [1989] "Motion and Structure from two Perspective Views: Algorithms, Error Analysis and Error Estimation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, (5), pp. 451-476, May 1989.
- Zhang, Z. Deriche, R. Faugeras, O. Luong, Q. [1996] "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry", INRIA Research Report 2273, May.
- Zhang, Z. [1998]"Determining the Epipolar Geometry and its Uncertainty: A Review", International Journal Of Computer Vision Vol. 27, (2), pp. 161-195. March/April 1998.
- Zhuang, X. e Huang T. [1986] "Two-view motion analysis: a unified algorithm" Journal Opt. Soc. America A, Vol. 3, 9, Sept. Pag. 1492-1500.