Pszeudo-műhold repülőgépek képifeldolgozása

Mályusz Attila Edmund
2021.12.2

A feladat a következő: Van egy 20 kilométer magasan több hónapon keresztül folyamatosan repülő a földről fényképeket készítő és feldolgozó repülőgép. A mi modellünkben a szárnyak modulokból épülnek fel, majd ezeken szárnyakon elhelyezünk egy-egy föld irányába néző kamerát. Egy haramadik kamera a repülőgép törzsnék a végén helyezkedik el és a két szárnyat látja. Az alap feladatunk az, hogy a kamerák által látott képekből szeretnék a lehető legjobb felbontású képet előállítani (Super-resolution). Ehhez segítséget nyújt az, ha tudjuk, hogy hol helyezkednek el egymáshoz képest a szárnyakon lévő kamerák. A kamerák helyének meghatározásában a következő adatokat fogjuk, majd igénybe venni: a két kamera által látott képek, a harmadik kamera által látott képek alapján és végűl a szárnyakon elhelyezett szenzorokból. A felhasznált algoritmusoknál még arra is figyelmet kell fordítanunk, hogy mivel csak a repülőgépen elérhető forrásokból számolunk és azonnal kellenek az adatok, így gyorsan futó algoritmusokat használhatunk csak. Ezzel feladattal már a Facebook is foglalkozott [1]. Ezt a projektet 2018-ban leállítottak. Végeredményben a repülő akkora volt, mint egy Boeing 737 melynek a szárnyfesztávolsága 34 méter, ellenben csak 400 kg volt a tömege, míg egy Boeing 737-nak 42 tonna.

A kamera mozgást leírhatjuk egy $g_t : \mathbb{R}^3 \to \mathbb{R}^3$ $t \in [0, T]$ függvényel. A g_t függvényre még a következőket követeljük meg

- $|g_t(v)| = |v|$ $\forall v \in \mathbb{R}^3$
- $g_t(u) \times g_t(v) = g_t(u \times v)$ $\forall u, v \in \mathbb{R}^3$

Azaz tartsa a távolságokat és a szögeket. Ebből a definícióból azt mutathatjuk meg [3] [6], hogy a kamera mozgást leíró transzformációk $SE(3) = \{g = (R, T) | R \in SO(3), T \in \mathbb{R}^3\}$ alakban állnak elő, ahol R a forgatást jelenti T pedig az eltolást. Megmutatható, hogy az $R \in SO(3)$ forgatás megközleíthető skew szimmetrikus mátrixokkal. A megközelítés lényege az, hogy az eredeti 9 változó és $det(R) = 1$ nehezen ellenőrizhető feltételek helyett 3 változóra kell optimalizálni. Ugyanígy a $g \in SE(3)$ elemei is megközelíthetőek.

A super-resolution feladatban kapunk n db gyenge felbontású képet (későbbiekben LR-k (low resolution)) egy adott jó felbontású képről HR. A cél az, hogy ezen LR képekből előállítsuk egy jó approximációját az eredeti HR képnek. A problémát mátrixokkal a következőképpen fogalmazhatjuk meg [4] [7]. Legyenek a kapott LR képek Y_k a keresett HR kép pedig X. Az LR képek készítése közben felmerülhetnek zajok V_k melyek Gauss eloszlásúak, a képek nem pontosan az eredeti képet tartalmazzák, így van egy eltolás operátor F_k, a mozgás során elmosódás is felép C és végül pedig a pixel szám kisebb lesz az LR képeken, így van egy down-sampling operator D. Így a következő relációt kapjuk az LR képek és HR kép között $Y_k = DCF_kX + V_k$. Feltesszük, hogy D-t, C-t és F_k-t tudjuk és C eltolás invariáns.

A X megtalálására sok algoritmus született. A nem neurális hálós technikák közül a [9] összefoglaló cikk szerint a wavelet interpolációs technika bizonyult a legjobbnak, egyrészt sokkal jobb eredményeket produkált másrészt gyorsabb is volt, mint a többi algoritmus. A wavelet interpolation a wavelet függvényeken alapszik [5]. A wavelet függvényeket két lépésben készítjük. Először veszünk egy $\varphi(x) \in L^2(\mathbb{R})$ skála függvényt, ennek a függvénynek a következő eltolt és skálázott verzióival alakítunk ki altereket $\varphi(x)_{j,k} = 2^{\frac{j}{2}}\varphi(2^jx-k)$ $k, j \in \mathbb{Z}$. Legyen $V_j = \overline{Span_k(\varphi_{j,k}(x))}$. A választott skála függvénynek teljesítenie kell a következő feltételeket: a skála függvény ortogonális az egész eltoltjaira, $\forall j \in \mathbb{Z}, V_j$-re $\subset V_{j+1}$, ha $f \in L^2(\mathbb{R})$ és $\forall j \in \mathbb{Z}, f \in V_j$, akkor $f \equiv 0$, és végűl $V_\infty = L^2(\mathbb{R})$.
Ekkor tudunk választani egy $\psi(x)$ függvényt melynek ugyanúgy a binárisan skálázott és egész eltoltjaiból alkotunk altereket $W_j = \overline{Span_k(\psi_{j,k}(x))}$ és teljesül, hogy $\forall j \in \mathbb{Z}$ $V_{j+1} = V_j \oplus W_j$. Ekkor $\forall j \in \mathbb{Z}$ $V_j \oplus W_j \oplus W_{j+1} \oplus \cdots = L^2(\mathbb{R})$.
Ez pedig mutatja a wavelet függvények approximáció elméleti hasznosságát mégpedig, hogy adott $f \in L^2(\mathbb{R})$ függyvényre egy durva becslést adunk a V_j-vel, majd azt javítjuk a W_k-ból vett elemekkel. Azaz $f(t) = \sum_{k\in\mathbb{Z}} a_{J,K}\varphi_{J,k}(t) + \sum_{j\geq J}\sum_{k\in\mathbb{Z}} b_{j,k}\psi_{j,k}(t)$, ahol $a_{J,K} = \int f(t)\varphi_{J,k}(t)dt$ és $b_{j,k} = \int f(t)\psi_{J,k}(t)dt$.

Képekre azaz a $f \in L^2(\mathbb{R}^2)$ esetre megmutatható, hogy $V_j^2 = V_j \otimes V_j$ [2] [8] módon képzett alterek kielégítik a skálára vontakozó feltételeket és a $2D$-s skála függvény előáll a következőképpen $\Phi(s,t) = \varphi(s)\varphi(t)$. A $2D$-s wavelet függvények V_j^2-ben, pedig a horizontális, vertikális és átlós irányban tárolnak információt.
Az interpolációt az $1D$-s esetben úgy kezeljük, hogy legyen $f \in V_0$ és $J \leq -1$, ekkor $V_0 = V_J \oplus \bigoplus_{j=J}^{-1} W_j$. Az f-nek megkapjuk M helyen az értéket. Ekkor kihasználjuk, hogy f felirható az előző alterek lineáris kombinációjaként és azt, hogy a szummák a lineáris kombinációban nem tartalmaznak végtelen sok elemet. Így most már mátrixos alakban fel lehet írni a feladatot $f = G_Ja_J + \sum_{j=J}^{-1} H_jb_j$, ahol $f = (f(t_i))_{i=0,\ldots,P-1}$, $a_J = (a_{J,k})_{k\in S_J}$, $b_j = (b_{j,k}k \in S_j)$, $G_J = (\varphi_{K,k}(t_i))_{i=0,\ldots,P-1}^{k\in S_j}$, $H_j = ((\psi_{j,k}(t_i)))_{i=0,\ldots,P-1}^{k\in S_j}$.
Ekkor először a_J értékeket becsüljük meg $\hat{a}_J$ regularized least squares módszerrel, mégpedig $f \approx G_Ja_J$-ból. Majd b_J-t becsüljük meg $g_J = f - G_J\hat{a}_J$-ből. Ha több M-es mintánk is van, akkor $f^{(i)} = G_J^{(i)}a_J$-ra kell az előző módszert elvégeznünk.
$2D$-s esetben ugyanezt kell csinálnunk, viszont a következő módosítással $f \approx (G_{J_s} \otimes G_{J_t})a_J$, ahol $\otimes$ a Kronecker szorzat és $G_{J_s} \otimes G_{J_t}$ megint a skálához tartozó eggyütthatókat jelenti.

Irodalomjegyzék

[1] Facebook aquila. `https://en.wikipedia.org/wiki/Facebook_Aquila`.

[2] Tensor product. `https://www.quantiki.org/wiki/tensor-product`.

[3] P. D. Cremers. Moving Camera. `https://vision.in.tum.de/_media/teaching/ss2019/mvg2019/material/multiviewgeometry2.pdf`.

[4] M. Elad and A. Feuer. Super-resolution reconstruction of an image. In *Proceedings of 19th Convention of Electrical and Electronics Engineers in Israel*, pages 391–394, 1996.

[5] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., USA, 2006.

[6] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.

[7] P. Milanfar. *Super-Resolution Imaging*. CRC Press, 2011.

[8] N. Nguyen and P. Milanfar. A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution). *Circuits, Systems and Signal Processing*, 19(4):321–338, Jul 2000.

[9] D. Yang, Z. Li, Y. Xia, and Z. Chen. Remote sensing image super-resolution: Challenges and approaches. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pages 196–200, 2015.