

Görüntü İşleme ile Mikro İnsansız Hava Araçlarının Algılanması

Fatih Gökçe, Sertaç Olgunsoylu, Göktürk Üçoluk, Erol Şahin, Sinan Kalkan

KOVAN Araştırma Laboratuvarı, Bilgisayar Mühendisliği Bölümü, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

Email: fgokce@ceng.metu.edu.tr, sertac.olgunsoylu@gmail.com, {ucoluk, erol, skalkan}@ceng.metu.edu.tr

Özetçe — Mikro insansız hava araçlarının (mİHA) algılanması, özellikle mİHA'ların diğer mİHA'ları navigasyon, ortak çalışma gibi amaçlarla algılamaları gereken çoklu mİHA kontrol senaryolarında kritik öneme sahiptir. Bu çalışmada, bilgisayarla görmeyi mİHA'ların çevreyi algılaması için kullanan çalışmalardan farklı olarak, birbirlerini algılamalarına odaklandık ve bilgisayarla görmenin bu amaçla kullanımının iyi bir alternatif olup olmadığını irdeledik. Bunun için, “kelime torbası” modelini bazı değişiklikler yaparak kullandık ve bilgisayarla görmede yaygın olan çeşitli nesne tanıma özelliklerini değerlendirerek bir mİHA'nın bu yöntemle algılanabileceğini gösterdik.

Anahtar Kelimeler—Görüntü işleme, mikro insansız hava aracı, algılama, kelime torbası.

I. GİRİŞ VE YÖNTEM

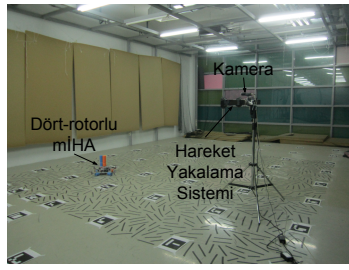
mİHA'lara olan ilgi geçtiğimiz birkaç on yılda oldukça artmıştır. Artan bu ilginin temel sebebi gelişen mikro-denetleyici ve sensör teknolojilerinin oldukça yetenekli platformları ucuz maliyetle üretmeye imkan sağlamasıdır. Uçuş özellikleri sebebiyle mİHA'lar gözetleme ve keşif gibi görevlerde kara araçlarına göre önemli avantajlara sahiptirler. mİHA'lara olan ilgiye paralel olarak bu platformlar, sürü robot bilimi çalışmalarında da kullanılmaya başlanmıştır [1]. Sürüdeki tüm mİHA'ların ayrı ayrı operatörler tarafından kontrol edilmesi çok zor olduğundan, sadece bir tanesinin kontrol edilmesi veya sürüye tek bir görev atanması öne çıkmaktadır. Bu durumda mİHA'ların birbirlerini algılamaları gerekmektedir.

mİHA'ların birbirlerini algılamaları için (1) küresel konumlama sistemi (GPS) [2], (2) kablosuz iletişim teknolojileri ile elde edilebilen radyo sinyallerinin havada uçuş süresi, birden fazla alıcıya ulaşma zaman farkları, geliş açısı kestirimi ve sinyal gücü gibi bilgilerin biri veya birkaçının kullanımı [3], (3) kızılötesi sinyaller [1] ve (4) ses sinyalleri [4] gibi çeşitli yöntemler kullanılabilir. Fakat bu yöntemler kullanılabilecekleri ortam ve mesafe ölçümündeki sağlayabilecekleri hassasiyet açısından kısıtlamalara neden olmaktadır.

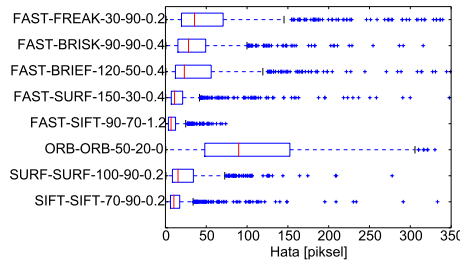
Nesne algılama ve tanımlama, bilgisayarla görme ve görüntü tanıma alanlarında oldukça çalışılmış; insan [5], taşıt algılama [6] ve manzara sınıflandırma [7] gibi oldukça başarılı uygulamalar gerçekleştirilmiştir. Bu uygulamalarda, öncelikle görüntü üzerindeki dikkat çekici kilit noktalar bulunarak, değişik transformasyonlar altında bile mümkün olduğunca ayırıcı özellikteki alanlar belirlenir. Boyuttan bağımsız özellik transformasyonu (SIFT) [8], hızlandırılmış gürbüz özellikler (SURF) [9], hızlı köşe sezimi (FAST) [10] ve yönlendirilmiş FAST-döndürülmüş ikili gürbüz bağımsız temel özellikler (ORB) [11] bu amaçla kullanılan başlıca yöntemlerdir. Kilit noktaların bulunmasından sonra bu noktaların yine değişik transformasyonlardan bağımsız şekilde tanımlanması gerekir.

Özellik tanımlama da denilen bu aşama için literatürde kullanılan yöntemlerin başlıcaları SIFT, SURF, FAST, ORB, ikili gürbüz bağımsız temel özellikler (BRIEF) [12], ikili gürbüz değişmeyen ölçeklenebilir kilit noktaları (BRISK) [13], hızlı retina kilit noktalarıdır (FREAK) [14]. Bulunan ve tanımlanan özellikler genellikle çok boyutludur ve bu durum özelliklerin dağılımlarının nesne algılama veya tanımlamada kullanımını güçleştirir. Bunu aşmak için literatürde popüler olarak kelime torbası (BOW) yöntemi kullanılmaktadır [15]. Bu yöntemde, eğitim aşamasında kullanılacak tüm nesnelere ait görüntülerdeki özellikler bir araya toplanarak, belirli sayıda kümeye (sözlük boyutu, k) ayrılırlar. Bu kümeler, nesnelere ait özellikler yerine, kümelere ait etiketler kullanılarak elde edilen özellik histogramları ile tanımlanmasında kullanılacaktır. Eğitim setinde bulunan nesnelere ait görüntülerden elde edilen histogramlar kullanılarak destek vektör makinası (SVM) [16] gibi bir sınıflandırıcı eğitilir. Test aşamasında, görüntü üzerinde bir pencere gezdirilerek, her bir pozisyonu için pencere içerisindeki özelliklere ait histogram çıkarılır ve eğitilmiş olan sınıflandırıcıya verilerek ilgili pencere içerisinde öğrenilen bir nesne bulunup bulunmadığı belirlenir. Burada, pencerenin boyutu algılanacak nesnenin boyutunu sınırlandırmaktadır. Farklı boyutlarda pencereler kullanmak bu problemi belli ölçüde giderebilir, fakat verimli bir çözüm elde edilmemiş olur.

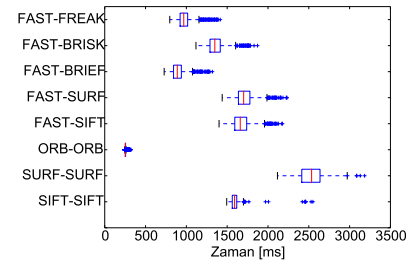
Çalışmamızda, mİHA'ları algılamak için yukarıda bahsedilen diğer yöntemlerin sahip oldukları kısıtlar düşünülerek, bilgisayarla görmenin uygun bir alternatif olarak kullanılıp kullanılmayacağını değerlendirdik. Bunun için BOW yöntemini bazı değişiklikler yaparak kullandık ve literatürdeki farklı kilit noktası bulma ve özellik tanımlama yöntemlerini karşılaştırdık. Normal BOW yönteminde kullanılan pencere yaklaşımı yerine, bulunan özelliklerin N en yakın komşusunu kullandık. Test aşamasında mİHA'nın bulunduğu noktayı belirlemek için, görüntü üzerinde bulunan her bir f özellik noktası için N komşuluğuna ait histogramları hesapladık. Bu histogramları önceden eğittiğimiz sınıflandırıcımıza vererek mİHA olarak sınıflandırılanları bulduk. mİHA olarak sınıflandırılan histogramlara karşılık gelen f noktalarının geometrik merkezi mİHA'nın bulunduğu noktadır. Amacımız gerçek zamanlı bir sistem geliştirmek değil, bilgisayarla görmenin iyi bir algılama performansı sağlayıp sağlayamayacağını değerlendirmektir. Yine de kamera görüntüsünün sadece bir kısmını işlemeyi yeterli kılacak farklı metodların entegrasyonu sayesinde gerçek zamanlı bir sisteme ulaşmak çok da uzak gözükmemektedir. Ayrıca, çalışmamız sadece bir tane mİHA'nın algılanmasını sağlamakta olup, çoklu mİHA algılama özellikleri henüz bulunmamaktadır.



(a)



(b)



(c)

Şekil 1. (a) Deney düzeneği. (b) Algılama performansları. Her bir kutu diyagramı, ilgili konfigürasyona ait en iyi sonuca karşılık gelmektedir. Konfigürasyonların yanındaki 3 rakam sırasıyla N , k ve T 'dir. Kutu diyagramında kutunun her iki ucu ve arasındaki kırmızı çizgi sırasıyla 1., 3. çeyrekler ve medyan değeri gösterir. Kutucukların sol ve sağdaki dikey çizgiler ise en küçük ve en büyük aykırı olmayan değeri gösterir. Kutu içinde kalan kısım %50, dikey çizgiler arasında kalan kısım ise %95 güven aralığındadır. Aykırı değerler + ile gösterilmiştir. (c) $N = 150$, $k = 100$ ve $T = 0$ için çalışma zamanı performansları.

II. DENEYLER VE SONUÇLAR

Deneylerimizde, özel tasarlanmış, aynı eksenlerdeki motorları arası mesafe 60 cm olan dört-rotorlu mİHA'yı $6m \times 10m \times 3.2m$ 'lik bir test ortamında (Şekil 1(a)) uçurarak Basler Scout marka bir kamera ile 1032×778 çözünürlükte, 30 fps hızda, gri tonlamalı görüntüler kaydettik. mİHA'nın gerçek pozisyonunu VisualeyezTM II VZ4000 hareket yakalama sistemi kullanarak elde ettik. Görüntülerde mİHA kameraya yaklaşma-uzaklaşma, yatayda konum değiştirme, yükselip-alçalma, kendi etrafında dönme gibi hareketleri yapmaktadır ve toplam 860 kare görüntü bulunmaktadır. Eğitim aşaması için farklı boyutlarda 99 adet mİHA'ya, 100 adet de ortama ait görüntü kullandık.

Testlerimizde farklı kilit noktası bulma ve özellik tanımlama yöntemlerinin algılama performansına olan etkisini üç farklı parametrenin çeşitli değerleri için inceledik. Parametreler ve değişim aralıkları şöyledir: (1) BOW yönteminde kullanılan *en yakın komşu sayısı*, N (10'dan 100'e 10'luk aralıklarla), (2) *sözlük boyutu*, k (10'dan 150'ye 10'luk aralıklarla) ve (3) SVM sınıflandırıcıya ait güven seviyesini sınırlandırmak için kullanılan *eşik değeri*, T (0'dan 2'ye 0.2'lik aralıklarla). Her bir kilit noktası bulma ve özellik tanımlama yöntemi çifti için, mİHA'yı tüm karelerde ve en az hata ile bulmayı sağlayan parametre setini tespit ettik. Hata ölçütü olarak 2-boyutlu resim düzleminde yöntemimizin bulduğu noktanın mİHA'nın gerçek bulunduğu noktaya uzaklığını kullandık. Şekil 1(b)'de her bir yöntem çiftinin, en iyi performansa sahip parametre seti için başarımlarını gösterilmiştir. En iyi başarımları FAST-SIFT çifti sağlamıştır. SIFT, SURF ve FAST kilit noktası bulma yöntemleri de boyuttan bağımsız çalıştıkları için oldukça iyi sonuç vermişlerdir. FAST'ın yönelim ve BRIEF'in dönme eklentileriyle birleştirildiği ORB yöntemi en kötü sonucu vermiştir. Bu, ORB'daki eklentilerin problemimiz için çok uygun olmadığını göstermektedir.

Şekil 1(c), Intel Xeon E5430 işlemci kullanılarak yapılan çalışma zamanı sonuçlarını içermektedir. İkili özellikler kullanıldığı için ORB, daha hızlı çalışmaktadır. Gerçek zamanlı bir çalışma için yöntemimizin iyileştirilmesi gerekmektedir. Aynı görüntüde bulunan özelliklerin sayıca fazla olması sebebiyle SURF-SURF, SIFT-SIFT'e göre beklenenin aksine daha kötü performans göstermiştir. FAST, SIFT ve SURF'e göre genelde daha iyi performans göstermekle birlikte beraber kullanıldığı özellik tanımlama yönteminin de etkisi gözlenmektedir.

TEŞEKKÜR

Fatih Gökçe, Süleyman Demirel Üniv. adına ÖYP Programı'nda kayıtlı bulunmaktadır. Çalışmada kullanılan Yüksek

Başarımli Hesaplama kaynakları, ODTÜ Bilgisayar Mühendisliği Bölümü tarafından sağlanmıştır.

KAYNAKÇA

- [1] J. Roberts, T. Stirling, J. Zufferey, and D. Floreano, "3-d relative positioning sensor for indoor flying robots," *Autonomous Robots*, 2012.
- [2] C. Virágh, G. Vársárhelyi, N. Tarcai, T. Szórényi, G. Somorjai, T. Nepusz, and T. Vicsek, "Flocking algorithm for autonomous flying robots," *Bioinspiration & Biomimetics*, vol. 9, no. 2, p. 025012, 2014.
- [3] G. Mao, B. Fidan, and B. D. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, no. 10, pp. 2529–2553, 2007.
- [4] M. Basiri, F. Schill, D. Floreano, and P. Lima, "Audio-based relative positioning system for multiple micro air vehicle systems," in *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 886–893, 2005.
- [6] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Tran. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.
- [7] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [10] M. Trajnkovic and M. Hedley, "Fast corner detection," *Image and Vision Computing*, vol. 16, no. 2, pp. 75–87, 1998.
- [11] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: An efficient alternative to SIFT or SURF," *Int. Conf. on Computer Vision*, pp. 2564–2571, 2011.
- [12] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," *European Conf. on Computer Vision*, vol. 6314, pp. 778–792, 2010.
- [13] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," *Int. Conf. on Computer Vision*, pp. 2548–2555, 2011.
- [14] P. Vanderghenst, R. Ortiz, and A. Alahi, "FREAK: Fast retina key-point," *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 0, pp. 510–517, 2012.
- [15] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," *Workshop on Statistical Learning in Computer Vision*, pp. 1–22, 2004.
- [16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.