Are recent advances in automated benthic macroinvertebrate taxa identification a viable option to manual keying?

Kristian Meissner

Finnish Environment Institute SYKE +
Ärje[1], Kärkkäinen[1], Tirronen[2], Turpeinen[3], Juhola[4], Joutsijoki[4], Raitoharju[5], Gabbouj[5] , Kiranyaz[5,6]

University of Jyväskylä:
- Dept. of Mathematics and Statistics[1]
- Dept. of Mathematical Information Technology[2]
- Department of Physics[3],

University of Tampere:
- School of Information Sciences[4]

Tampere University of Technology:
- Department of Signal Processing [5]

Qatar University[6]
- Department of Electrical Engineering

Research program for environmental data aquisition and use

Why use automated taxa identification?

- growing demand for biological water quality assessments
- mismatch between resources and demands
- identification accounts for 1/3 of the cost of each macroinvertebrate sample (in Finland)

Possible solution:

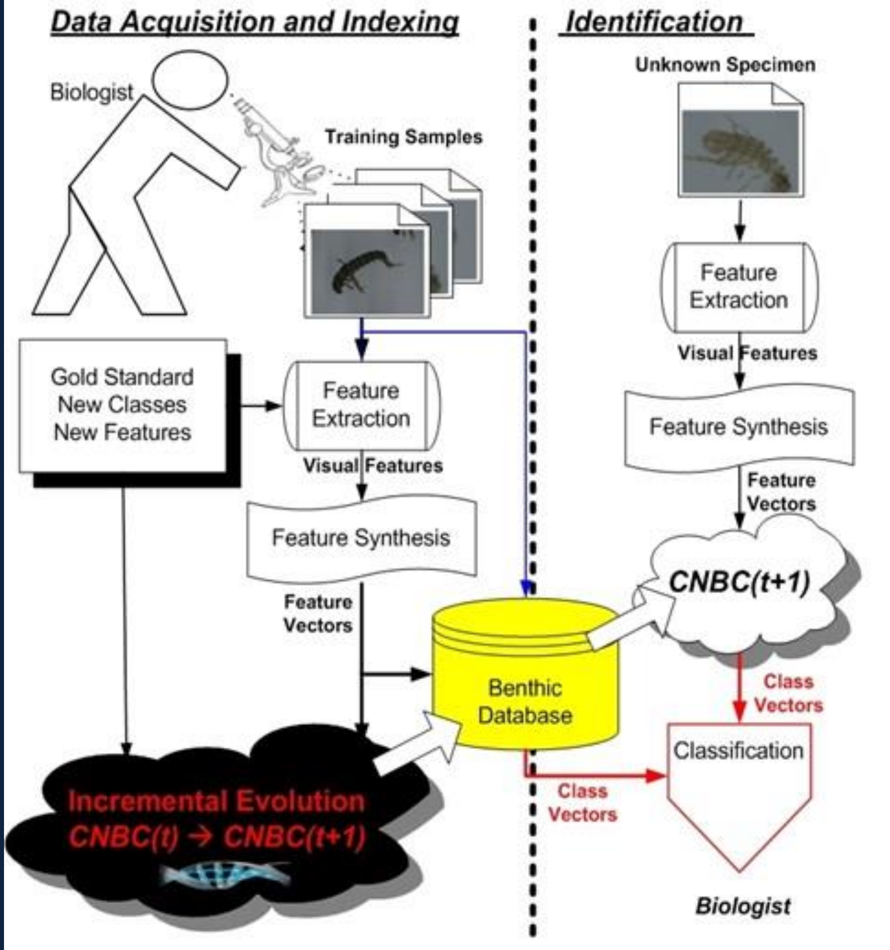
Cut identification cost with computer vision

- cheap
- Fast, allows higher sample volumes

Research program for environmental data aquisition and use

SYKE

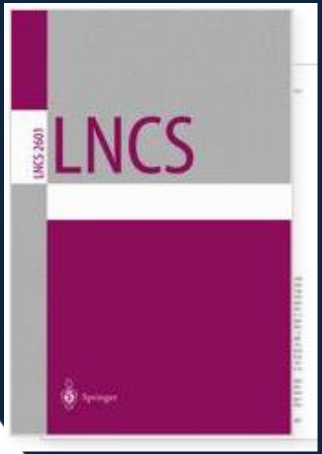
How automated taxa identification works



Research program for environmental data aquisition and use

First trials: Prototype 1

8 taxa set, flatbed scanner, 1592 single posture images



LNCS

Springer

Results:
80% correct and 400-10000 times faster than a human expert (Tirronen et al. 2009)

Computers in Biology and Medicine

An International Journal

Results:
97% correct (Kiranyaz et al. 2011)

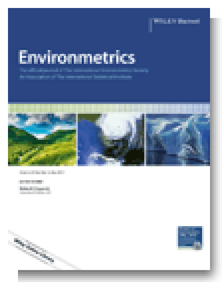
Research program for environmental data aquisition and use

SYKE

Additional studies

(8->) 35 species, 6814 single posture images, flatbed scanner

- Results:
- New classifier
- Random Bayes Array (RBA)
- 81% correct
- RBA did not use its full potential (naive priors)
- Ärje et al. 2013, Environmetrics


Environmetrics

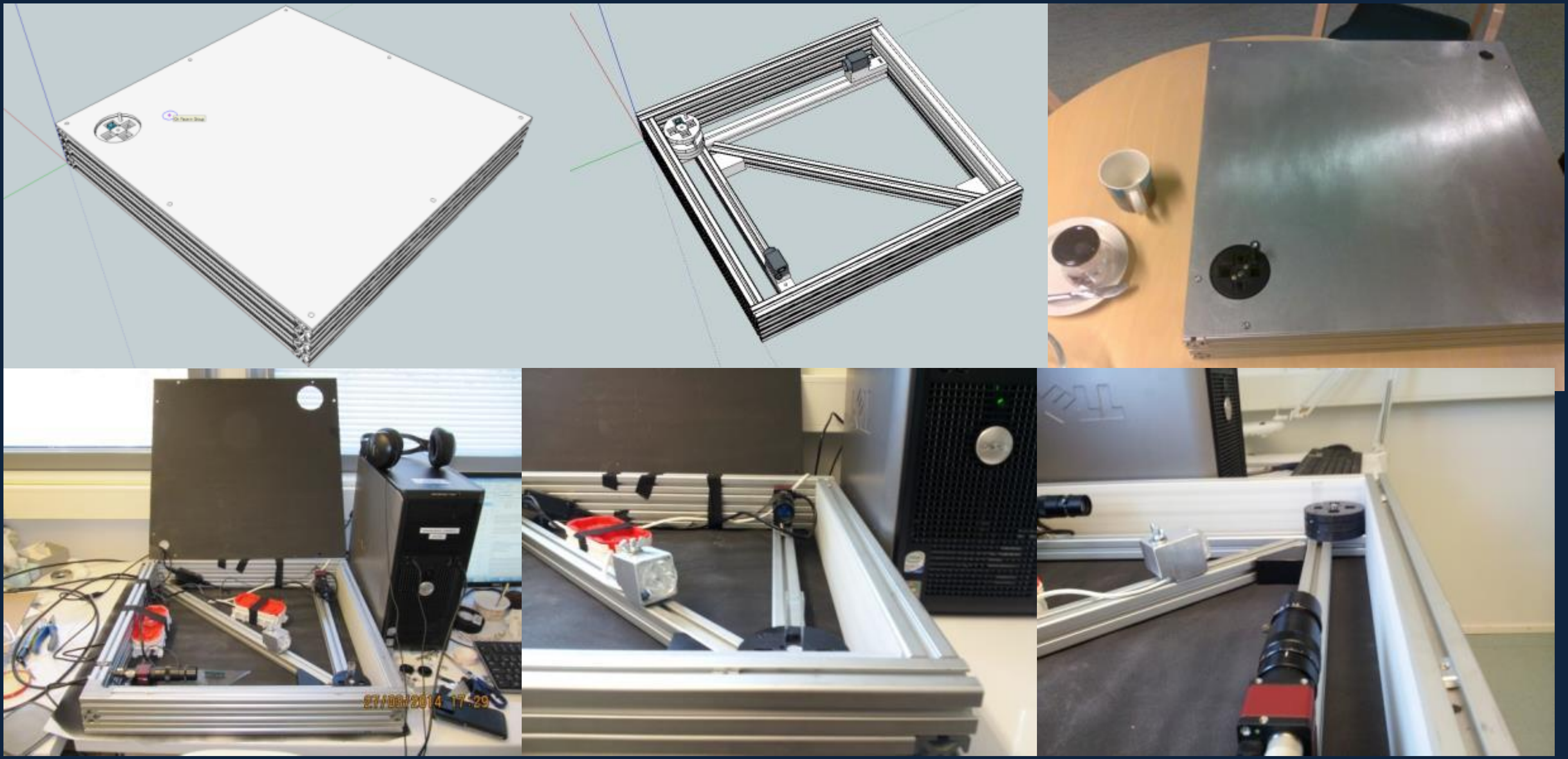
- Prototype1: too work intensive =>
- New prototype development started in 2013
- Prototype 2: an extension of the picking process

- New feature:
- Two posture imaging => two independent identification results / specimen = better identification

Research program for environmental data aquisition and use

SYKE

Prototype 2: two posture design



Research program for environmental data aquisition and use

SYKE

Prototype 2

Most complex data set to date

  – consists of 42 species, 6670 images

- Two postures from each individual
- Half of the data used for training the algorithm, half for testing

Prototype 2: Actual footage

Left camera


Baetis muticus


Baetis rhodani


Baetis rhodani 2


Limnius volckmari


Simuliidae


Oulimnius tuberculatus

Right camera


Baetis muticus


Baetis rhodani


Baetis rhodani 2


Limnius volckmari


Simuliidae


Oulimnius tuberculatus

How the RBA algorithm reaches a decision

Research program for environmental data aquisition and use

Limnius volckmari

Limnius volckmari

Baetis rhodani	Baetis muticus	Limnius volckmari	Simuliidae	Oulimnius tub.	Final posterior
0.07	0.05	0.88	0	0	0.8
0.10	0.08	0.72	0	0	

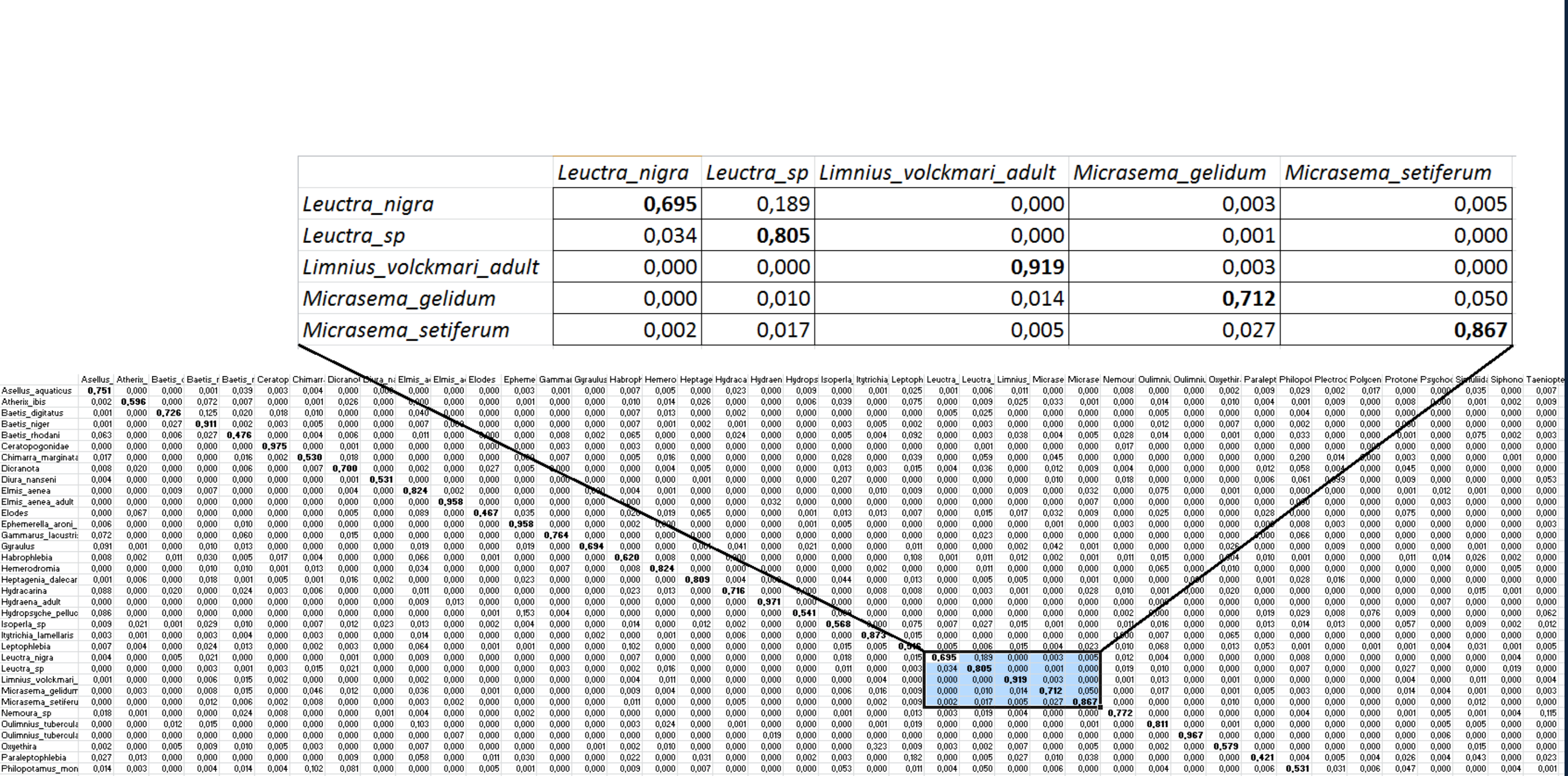
Results :
Overall Classification errors (CE)

	Right camera	Left camera	Both cameras
Mean(CE)	0.27	0.29	**0.22**
Sd(CE)	0.0015	0.0016	0.0012

Two postures clearly improve identification:
From 71% -> 78% correct

Research program for environmental data aquisition and use

Confusion matrix



	Leuctra_nigra	*Leuctra_sp*	*Limnius_volckmari_adult*	*Micrasema_gelidum*	*Micrasema_setiferum*
Leuctra_nigra	**0,695**	0,189	0,000	0,003	0,005
Leuctra_sp	0,034	**0,805**	0,000	0,001	0,000
Limnius_volckmari_adult	0,000	0,000	**0,919**	0,003	0,000
Micrasema_gelidum	0,000	0,010	0,014	**0,712**	0,050
Micrasema_setiferum	0,002	0,017	0,005	0,027	**0,867**

	Asellus_	Atherix_	Baetis_	Baetis_r	Baetis_r	Ceratop	Chimarr	Dicrano	Diura_na	Elmis_a	Elmis_a	Elodes	Epheme	Gammar	Gyraulus	Habroph	Hemero	Heptage	Hydraca	Hydraen	Hydrops	Isoperla_	Itytrichia	Leptoph	Leuctra_	Leuctra_	Limnius_	Micrase	Micrase	Nemour	Oulimniu	Oulimniu	Oxyethir	Paralept	Philopot	Plectroc	Polycen	Protone	Psychoc	Simuliida	Siphono	Taeniopte
Asellus_aquaticus	**0,751**	0,000	0,000	0,001	0,039	0,003	0,004	0,000	0,000	0,000	0,000	0,000	0,003	0,001	0,000	0,007	0,005	0,000	0,023	0,000	0,009	0,000	0,001	0,025	0,001	0,006	0,011	0,000	0,000	0,008	0,000	0,000	0,002	0,009	0,029	0,002	0,017	0,000	0,000	0,035	0,000	0,007
Atherix_ibis	0,002	**0,596**	0,000	0,072	0,007	0,000	0,001	0,026	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,010	0,014	0,026	0,000	0,000	0,006	0,039	0,000	0,075	0,000	0,009	0,025	0,033	0,001	0,000	0,014	0,000	0,010	0,004	0,001	0,009	0,000	0,008	0,000	0,001	0,002	0,009
Baetis_digitatus	0,001	0,000	**0,726**	0,125	0,020	0,018	0,010	0,000	0,000	0,040	0,000	0,000	0,000	0,000	0,000	0,007	0,013	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,005	0,025	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Baetis_niger	0,001	0,000	0,027	**0,911**	0,002	0,003	0,005	0,000	0,000	0,007	0,000	0,000	0,000	0,000	0,000	0,007	0,001	0,002	0,001	0,000	0,000	0,003	0,005	0,002	0,000	0,003	0,000	0,000	0,000	0,000	0,012	0,000	0,007	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Baetis_rhodani	0,063	0,000	0,006	0,027	**0,476**	0,000	0,004	0,006	0,000	0,011	0,000	0,000	0,000	0,008	0,002	0,065	0,000	0,000	0,024	0,000	0,000	0,005	0,004	0,092	0,000	0,003	0,038	0,004	0,005	0,028	0,014	0,000	0,001	0,000	0,033	0,000	0,000	0,001	0,000	0,075	0,002	0,003
Ceratopogonidae	0,000	0,000	0,000	0,000	0,000	**0,975**	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,003	0,000	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,017	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Chimarra_marginata	0,017	0,000	0,000	0,000	0,016	0,002	**0,530**	0,018	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,005	0,016	0,000	0,000	0,000	0,000	0,028	0,000	0,039	0,000	0,059	0,000	0,045	0,000	0,000	0,000	0,000	0,000	0,000	0,200	0,014	0,000	0,003	0,000	0,000	0,001	0,000
Dicranota	0,008	0,020	0,000	0,000	0,006	0,000	0,007	**0,700**	0,000	0,002	0,000	0,027	0,005	0,000	0,000	0,000	0,004	0,005	0,000	0,000	0,000	0,013	0,003	0,015	0,004	0,036	0,000	0,012	0,009	0,004	0,000	0,000	0,000	0,012	0,058	0,004	0,000	0,045	0,000	0,000	0,000	0,000
Diura_nanseni	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,001	**0,531**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,207	0,000	0,000	0,000	0,000	0,000	0,010	0,000	0,018	0,000	0,000	0,000	0,006	0,061	[illegible]	0,000	0,009	0,000	0,000	0,000	0,053
Elmis_aenea	0,000	0,000	0,009	0,007	0,000	0,000	0,000	0,004	0,000	**0,824**	0,002	0,000	0,000	0,000	0,000	0,004	0,001	0,000	0,000	0,000	0,000	0,000	0,010	0,009	0,000	0,000	0,009	0,000	0,032	0,000	0,075	0,000	0,001	0,000	0,000	0,000	0,000	0,001	0,012	0,001	0,000	0,000
Elmis_aenea_adult	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	**0,958**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,032	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,003	0,000	0,000	0,000
Elodes	0,000	0,067	0,000	0,000	0,000	0,000	0,000	0,005	0,000	0,089	0,000	**0,467**	0,035	0,000	0,000	0,020	0,019	0,065	0,000	0,000	0,001	0,013	0,013	0,007	0,000	0,015	0,017	0,032	0,009	0,000	0,025	0,000	0,000	0,028	0,000	0,000	0,000	0,075	0,000	0,000	0,000	0,000
Ephemerella_aroni_	0,006	0,000	0,000	0,000	0,010	0,000	0,000	0,000	0,000	0,000	0,000	0,000	**0,958**	0,000	0,000	0,002	0,000	0,000	0,000	0,000	0,001	0,005	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,003	0,000	0,000	0,000	[illegible]	0,008	0,003	0,000	0,000	0,000	0,000	0,000	0,003
Gammarus_lacustri	0,072	0,000	0,000	0,000	0,060	0,000	0,000	0,015	0,000	0,000	0,000	0,000	0,000	**0,764**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,023	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,066	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Gyraulus	0,091	0,001	0,000	0,010	0,013	0,000	0,000	0,000	0,000	0,019	0,000	0,000	0,019	0,000	**0,694**	0,000	0,000	0,001	0,041	0,000	0,021	0,000	0,000	0,011	0,000	0,000	0,002	0,042	0,001	0,000	0,000	0,000	0,026	0,000	0,000	0,009	0,000	0,000	0,000	0,001	0,000	0,000
Habrophlebia	0,008	0,002	0,011	0,030	0,005	0,017	0,004	0,000	0,000	0,066	0,000	0,001	0,000	0,000	0,000	**0,620**	0,008	0,000	0,000	0,000	0,000	0,000	0,000	0,108	0,001	0,011	0,012	0,002	0,001	0,011	0,015	0,000	0,004	0,010	0,001	0,000	0,000	0,011	0,014	0,026	0,002	0,000
Hemerodromia	0,000	0,000	0,000	0,010	0,010	0,001	0,013	0,000	0,000	0,034	0,000	0,000	0,000	0,007	0,000	0,008	**0,824**	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,011	0,000	0,000	0,000	0,000	0,065	0,000	0,010	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,005	0,000
Heptagenia_dalecar	0,001	0,006	0,000	0,018	0,001	0,005	0,001	0,016	0,002	0,000	0,000	0,000	0,023	0,000	0,000	0,000	0,000	**0,809**	0,004	0,000	0,000	0,044	0,000	0,013	0,000	0,005	0,005	0,000	0,001	0,000	0,000	0,000	0,000	0,001	0,028	0,016	0,000	0,000	0,000	0,000	0,000	0,000
Hydracarina	0,088	0,000	0,020	0,000	0,024	0,003	0,006	0,000	0,000	0,011	0,000	0,000	0,000	0,000	0,000	0,023	0,013	0,000	**0,716**	0,000	0,000	0,000	0,008	0,008	0,000	0,003	0,001	0,000	0,028	0,010	0,001	0,000	0,020	0,000	0,000	0,000	0,000	0,000	0,000	0,015	0,001	0,000
Hydraena_adult	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,009	0,013	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	**0,971**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,000	0,000
Hydropsyche_pelluc	0,086	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,153	0,004	0,000	0,000	0,000	0,000	0,000	0,000	**0,541**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,019	0,029	0,008	0,076	0,009	0,000	0,000	0,000	0,062
Isoperla_sp	0,009	0,021	0,001	0,029	0,010	0,000	0,007	0,012	0,023	0,013	0,000	0,002	0,004	0,000	0,000	0,014	0,000	0,012	0,002	0,000	0,000	**0,568**	0,000	0,075	0,007	0,027	0,015	0,001	0,000	0,011	0,016	0,000	0,000	0,013	0,014	0,013	0,000	0,057	0,000	0,009	0,002	0,012
Itytrichia_lamellaris	0,003	0,001	0,000	0,003	0,004	0,000	0,003	0,000	0,000	0,014	0,000	0,000	0,000	0,000	0,002	0,000	0,001	0,000	0,006	0,000	0,000	0,000	**0,873**	0,015	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,065	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Leptophlebia	0,007	0,004	0,000	0,024	0,013	0,000	0,002	0,003	0,000	0,064	0,000	0,001	0,001	0,000	0,000	0,102	0,000	0,000	0,000	0,000	0,000	0,015	0,005	[illegible]	0,005	0,006	0,015	0,004	0,023	0,010	0,068	0,000	0,013	0,053	0,001	0,000	0,001	0,001	0,004	0,031	0,001	0,005
Leuctra_nigra	0,004	0,000	0,005	0,021	0,000	0,000	0,000	0,001	0,000	0,009	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,000	0,000	0,000	0,000	0,018	0,000	0,015	**0,695**	0,189	0,000	0,003	0,005	0,012	0,004	0,000	0,000	0,000	0,008	0,000	0,000	0,000	0,000	0,000	0,004	0,000
Leuctra_sp	0,000	0,000	0,000	0,003	0,001	0,003	0,015	0,021	0,000	0,000	0,000	0,000	0,000	0,003	0,000	0,002	0,016	0,000	0,000	0,000	0,000	0,011	0,000	0,003	0,034	**0,805**	0,000	0,001	0,000	0,019	0,010	0,000	0,000	0,000	0,007	0,000	0,000	0,027	0,000	0,000	0,019	0,000
Limnius_volckmari_	0,001	0,000	0,000	0,006	0,015	0,002	0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,004	0,011	0,000	0,000	0,000	0,000	0,000	0,004	0,000	0,000	0,000	**0,919**	0,003	0,000	0,001	0,013	0,000	0,001	0,000	0,000	0,000	0,000	0,004	0,000	0,011	0,000	0,004
Micrasema_gelidum	0,000	0,003	0,000	0,008	0,015	0,000	0,046	0,012	0,000	0,036	0,000	0,001	0,000	0,000	0,000	0,009	0,004	0,000	0,000	0,000	0,000	0,006	0,016	0,009	0,000	0,010	0,014	**0,712**	0,050	0,000	0,017	0,000	0,001	0,005	0,003	0,000	0,000	0,014	0,004	0,001	0,000	0,003
Micrasema_setiferu	0,000	0,000	0,000	0,012	0,006	0,002	0,008	0,000	0,000	0,003	0,002	0,000	0,000	0,000	0,000	0,011	0,000	0,000	0,005	0,000	0,000	0,000	0,002	0,009	0,002	0,017	0,005	0,027	**0,867**	0,000	0,000	0,000	0,010	0,000	0,000	0,000	0,000	0,000	0,000	0,012	0,000	0,000
Nemoura_sp	0,018	0,001	0,000	0,000	0,024	0,008	0,000	0,000	0,001	0,004	0,000	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,013	0,003	0,019	0,004	0,000	0,000	**0,772**	0,000	0,000	0,000	0,000	0,004	0,000	0,000	0,001	0,005	0,001	0,004	0,115
Oulimnius_tubercula	0,000	0,000	0,012	0,015	0,000	0,000	0,000	0,000	0,000	0,103	0,000	0,000	0,000	0,000	0,000	0,003	0,024	0,000	0,001	0,000	0,000	0,000	0,001	0,019	0,000	0,000	0,000	0,000	0,001	0,000	**0,811**	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,005	0,005	0,000	0,000
Oulimnius_tubercula	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,019	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	**0,967**	0,000	0,000	0,000	0,000	0,000	0,000	0,006	0,000	0,000	0,000
Oxyethira	0,002	0,000	0,005	0,009	0,010	0,005	0,003	0,000	0,000	0,007	0,000	0,000	0,000	0,000	0,001	0,002	0,010	0,000	0,000	0,000	0,000	0,000	0,323	0,009	0,003	0,002	0,007	0,000	0,005	0,000	0,002	0,000	**0,579**	0,000	0,000	0,000	0,000	0,000	0,000	0,015	0,000	0,000
Paraleptophlebia	0,027	0,013	0,000	0,000	0,000	0,000	0,000	0,009	0,000	0,058	0,000	0,011	0,030	0,000	0,000	0,022	0,000	0,031	0,000	0,000	0,002	0,003	0,000	0,182	0,000	0,005	0,027	0,010	0,038	0,000	0,000	0,000	0,000	**0,421**	0,004	0,005	0,004	0,026	0,004	0,043	0,000	0,023
Philopotamus_mon	0,014	0,003	0,000	0,004	0,014	0,004	0,102	0,081	0,000	0,000	0,000	0,005	0,001	0,000	0,000	0,009	0,000	0,007	0,000	0,000	0,000	0,053	0,000	0,011	0,004	0,050	0,000	0,006	0,000	0,000	0,004	0,000	0,000	0,006	**0,531**	0,031	0,006	0,047	0,000	0,000	0,004	0,001
Plectrocnemia_con	0,000	0,002	0,000	0,000	0,000	0,000	0,001	0,031	0,015	0,000	0,000	0,000	0,003	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,009	0,058	0,000	0,001	0,000	0,001	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,085	**0,745**	0,017	0,015	0,000	0,000	0,010	0,000
Polycentropus_flav	0,107	0,000	0,000	0,015	0,029	0,000	0,000	0,002	0,001	0,000	0,000	0,000	0,024	0,000	0,000	0,001	0,000	0,003	0,000	0,000	0,151	0,007	0,000	0,000	0,004	0,001	0,000	0,000	0,000	0,014	0,000	0,000	0,000	0,000	0,053	0,101	**0,420**	0,001	0,000	0,000	0,000	0,065
Protonemura_sp	0,006	0,005	0,001	0,000	0,003	0,000	0,001	0,015	0,004	0,000	0,000	0,023	0,005	0,000	0,000	0,016	0,008	0,000	0,000	0,000	0,000	0,038	0,000	0,011	0,000	0,089	0,000	0,001	0,000	0,006	0,021	0,000	0,000	0,021	0,006	0,009	0,000	**0,651**	0,000	0,000	0,022	0,036
Psychodiidae	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,024	0,001	0,000	0,000	0,000	0,000	0,033	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,003	0,000	0,005	0,000	0,000	0,001	0,000	0,016	0,000	0,000	0,000	0,000	0,000	0,000	0,003	**0,914**	0,000	0,000	0,000
Simuliidae	0,019	0,000	0,000	0,007	0,006	0,000	0,000	0,002	0,000	0,045	0,000	0,000	0,010	0,000	0,000	0,038	0,000	0,000	0,009	0,000	0,000	0,003	0,004	0,016	0,000	0,000	0,031	0,004	0,002	0,000	0,002	0,000	0,024	0,011	0,003	0,000	0,000	0,000	0,006	**0,754**	0,000	0,002
Siphonoperla_burm	0,002	0,007	0,000	0,001	0,031	0,003	0,017	0,007	0,006	0,010	0,000	0,000	0,000	0,001	0,000	0,045	0,016	0,000	0,000	0,000	0,000	0,011	0,000	0,033	0,005	0,408	0,000	0,000	0,000	0,067	0,015	0,000	0,000	0,000	0,011	0,000	0,000	0,050	0,000	0,000	**0,255**	0,000
Taeniopteryx_nebul	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000	0,001	0,000	0,000	0,010	0,000	0,000	0,058	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,002	0,000	**0,919**

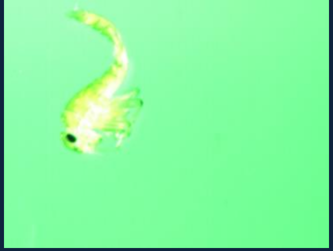
Tricky pictures

Research program for environmental data aquisition and use

Asellus aquaticus

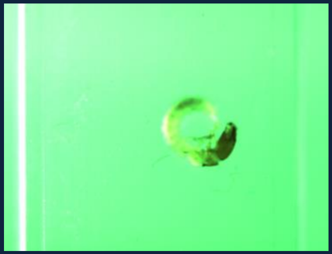
Baetis digitatus

Baetis niger



Baetis rhodani

Ceratopogonidae

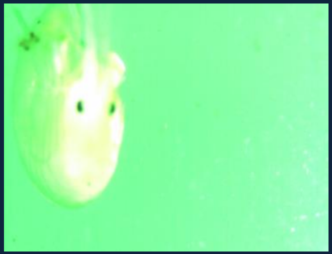


Chimarra marginata

Dicranota

Elmis aenea

Elodes



Gammarus lacustris

Habrophlebia

Leuctra nigra

Computer vision- a viable alternative to manual keying?

- Prototype 2 produced 78% correct with suboptimal pictures
- No priors – suboptimal performance
- Two postures improved the identification
- Sharper pictures (DOF, lighting) and more complex algorithms will increase the accuracy further

Benefits:

- Cheap (prototype 2 costs < 10000€)
- Functions with non expert operators
- Predictable error structure
- Additional metrics for free (size spectra / biomass)
- Prototype 3 under development, ready in spring of 2016 (Finnish Academy funding)
- First test of human vs. machine planned in 2016 (during human proficiency testing)
- Can be used solo or to augment DNA techniques
- Fully operational in under 4 years

Thank you!

Research program for environmental data aquisition and use

SYKE