Early Vision and Actions

Giulio Sandini
Istituto Italiano di Tecnologia
Robotics, Brain and Cognitive Sciences

Workshop on: Learning Data Representation: Hierarchies and Invariance
November 22-24, McGovern Institute for Brain Research, MIT

Not only vision in *computational vision*

Want to stress the role of vision in guiding and understanding human action and interaction…

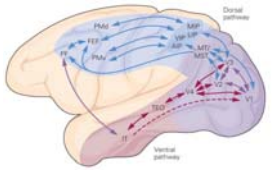


Figure 27-2 Cortical areas involved with intermediate-level visual processing. Many cortical areas in the macaque monkey, including V_1, V_2, V_3, V_4, and middle temporal area (**MT**), are involved with integrating local cues to construct contours and surfaces and segregating foreground from background. The shaded areas extend into the frontal and temporal lobes because cognitive output from these areas, including attention, expectation, and perceptual task, contribute to the process of scene segmentation. (**AIP**, anterior intraparietal cortex; **FEF**, frontal eye fields; **IT**, inferior temporal cortex; **LIP**, lateral intraparietal cortex; **MIP**, medial intraparietal cortex; **MST**, medial superior temporal cortex; **MT**, middle temporal cortex; **PF**, pre-frontal cortex; **PMd**, dorsal premotor

From: Kandell et.al.: Principles of Neural Sciences – 5th edition

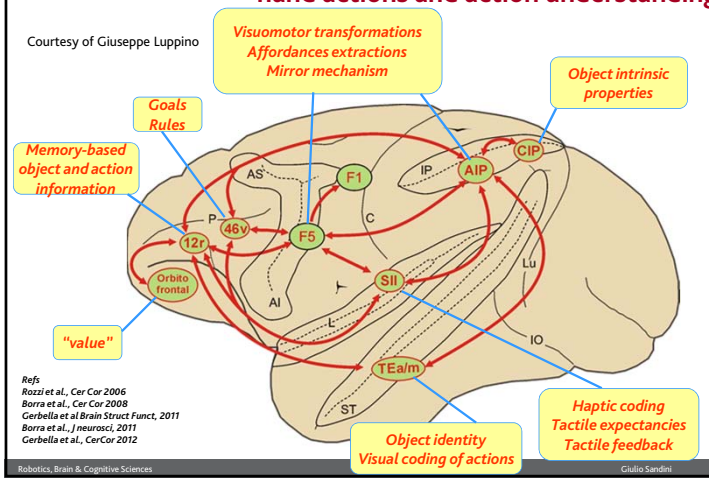
…the importance of integration with other sensory modalities (e.g. audition and haptics) and motor control & representation

Visual Observer Vs. Visual Actor

Most of "visual brain" is not "vision-only brain".

Cortical network centred in the AIP-F5 Circuit for hand actions and action understanding

Courtesy of Giuseppe Luppino



Refs
Rozzi et al., Cer Cor 2006
Borra et al., Cer Cor 2008
Gerbella et al Brain Struct Funct, 2011
Borra et al., J neurosci, 2011
Gerbella et al., CerCor 2012

e.g. visuo-somatic motor neurons in F4

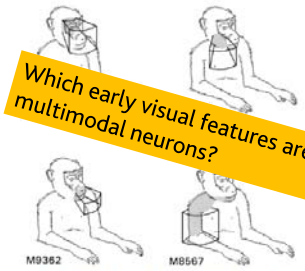


Fig. 4. Examples of different types of tactile and visual receptive fields of F4 bimodal neurons. Gray shaded area: tactile receptive fields. Solid around the tactile receptive fields: visual receptive fields.

Motor neurons in F4 have visual and tactile receptive field and are activated when:

1. The tactile RF is touched
2. An object enters into the corresponding visual RF
3. moved toward

Which early visual features are required to drive these multimodal neurons?

The visual R… "follows" the body segment!!

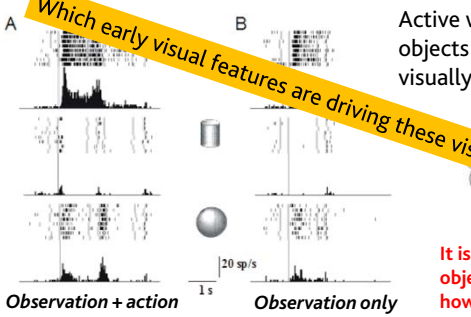
From: Fadiga, L., L. Fogassi, V. Gallese, and G. Rizzolatti, Visuomotor Neurons: ambiguity of the discharge or "motor" Perception? Internation Journal of Psychophysiology, 2000. 35: p. 165-177.

Object properties can also be coded in terms of motor primitives

- Position of objects can be coded in terms of the action required to reach it
- Trajectory of objects can be coded in terms of "collision trajectory" (which body part is going to hit)
- Shape/Size can be coded in terms of "grasp type" (small is whatever can be grasped with a pinch grasp)

…small like this…

…*canonical neurons* in F5 representing visual properties of objects in motor terms



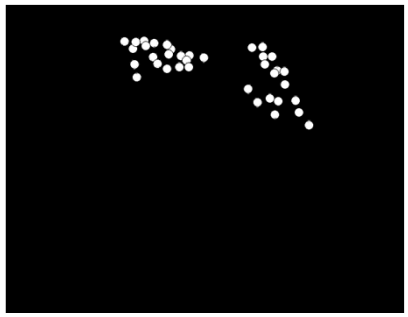
Which early visual features are driving these visuomotor neurons?

Active when manipulable objects are presented visually (grasp specific)

It is not the shape of the object that matters but how you can grasp it

From: Fadiga, L., L. Fogassi, V. Gallese, and G. Rizzolatti, Visuomotor Neurons: ambiguity of the discharge or "motor" Perception? Internation Journal of Psychophysiology, 2000. 35: p. 165-177.

Recognize an object without the object (shape from action)



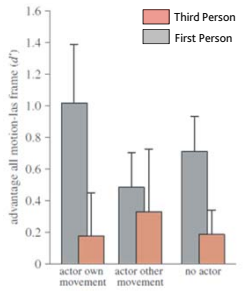
- How relevant is "action" in recognizing objects?
- Can the visual system exploit the view of haptic exploration to extract shape information of the explored object?

Study done with Francesco Campanella

Size information and reaching

human ability to extract some intrinsic object features, such as shape and size, from a visually presented reach-to-grasp movement.

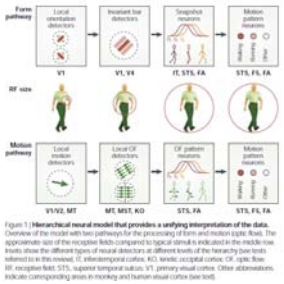
Size is coded in the action to reach (and point of view matters)



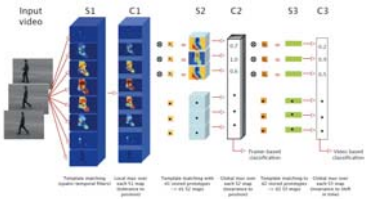
Campanella Sandini, Morrone, *Proc. R. Soc. B, 2010*

Modelling the Recognition of Biological Motion

Hierarchical models exploring the relationship between the ventral and the dorsal streams

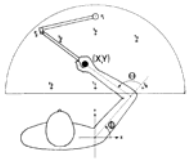


Giese & Poggio 2003

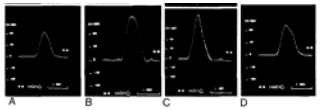


Jhuang , Serre, Wolf and Poggio- ICCV 2007

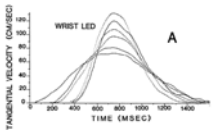
How can we exploit the regularities of human motion to "tune" early visual feature extraction?



Bell shaped velocity profile in reaching movements produce perspective invariant features on the retina of the actor and of the observer.



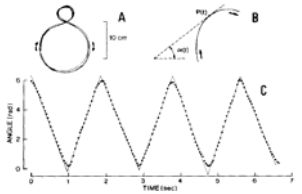
From: P. Morasso – Exp. Brain Res. - 1981



From: Atkeson & Hollerbach – J. Neuroscience 1985

Also: (Morasso 1981 , Morasso Mussa-Ivaldi 1982, Abend, Bizzi, Morasso, 1982)

Laquaniti's "2/3" Power Law



From: Laquaniti, Terzuolo, Viviani – Acta Psychologica - 1983

Isogony Principle: In drawing movements, equal angles are described in equal times (angular velocity tends to remain constant when the radius of curvature changes).
Viviani & Terzuolo 1982.

In rhythmic drawing movements, Laquaniti et. al. noted a power law relationship with proportionality constant k between the angular velocity $a(t)$ of the hand and the curvature of the trajectory path $c(t)$:

$$a(t) = k\, c(t)^{2/3}$$

"Human Motion Features" that could be exploited to *tune* early visual processes

- Regularity of velocity profile
- Position and time of changes of curvature (for segmentation)
- Relation between trajectory's curvature and velocity (e.g. 2/3 power low)
- Local rigidity (invariance of distance between joints)
- Common motion (not equal velocity but equal "visuo-motor pattern" – e.g. clouds of points with similar velocity profile or oscillations).
- Influence of gravity on trajectory.
- …

Questions for a possible discussion

- How can "motor knowledge" and the regularities of motor control be exploited to design/tune/learn how to compute early visual features?
- Is "looking at one's own body" a useful method to learn such features?
- How far can we go with only "bottom-up", feed forward models to explain human movement perception?
- What is innate and what is learned (and how)?

Thanks!