

II. Мета роботи

Метою роботи є аналіз проблем, що виникають при візуалізації тривимірних графічних об'єктів в реальному масштабі часу.

III. Проблеми візуалізації графічних об'єктів в реальному часі

У системах реального часу найважливішою характеристикою якості є тимчасова ефективність процесу відображення інформації в залежності від обсягу даних, що візуалізуються. Відображення інформації, звичайно являє собою тривимірну сцену, що складається з безлічі різних об'єктів. З метою підвищення якості цього типу програмного забезпечення, розробники йдуть на нетривіальні способи збільшення часової ефективності алгоритмів візуалізації стосовно конкретних відображуваних даними. Ці алгоритми базуються на методах комп'ютерної геометрії та машинної графіки. Їх завданнями є відсікання невидимих частин і відображення потрібних об'єктів оптимальним чином. Для виконання цих операцій використовуються спеціальні структури даних для представлення тривимірних сцен в пам'яті комп'ютера.

Загальна продуктивність системи візуалізації, як правило, складається із сукупності алгоритмів відсікання невидимих частин і оптимізації кінцевої моделі виведених об'єктів для остаточної візуалізації.

Найбільшими проблемами алгоритмів візуалізації тривимірних зображень в реальному часі є: відсікання невидимих частин, розрахунок освітленості, вибір методів зафарбовування об'єктів, представлення великих обсягів даних в пам'яті комп'ютера, вибір або побудова оптимального рівня деталізації для відображуваних об'єктів, забезпечення безперервності поверхонь для об'єктів з полігональною структурою, побудова оптимальної послідовності видимих об'єктів для ефективного використання апаратного прискорення, оптимізація програмних рішень під сучасні апаратні засоби.

Для представлення трьохвимірних зображень в пам'яті комп'ютера не розроблено однозначно ефективних структур даних, які б дозволили їх візуалізувати в реальному часі. Це пов'язано з труднощами досягнення компромісу між способом стиснення даних сцени і їх ефективною організацією для використання апаратного прискорення візуалізації. Насамперед, це зауваження відноситься до ландшафтів. Однак для певних типів графічних об'єктів розроблено комерційні рішення, але вони не дозволяють працювати з фіксованим рівнем деталізації.

Висновок

У роботі розглянуто основні проблеми візуалізації тривимірних графічних об'єктів в реальному часі.

Список використаних джерел

1. Банковский Ю.М., Галактионов В.А. О некоторых фундаментальных проблемах компьютерной (машинной) графики // Информационные технологии и вычислительные системы, 2004. - т. 4, с. 3-24

УДК 004.8

ПОБУДОВА СИСТЕМИ РОЗПІЗНАВАННЯ ЗВУКІВ ЛЮДСЬКОЇ МОВИ ЗА ДОПОМОГОЮ МЕЛ-ЧАСТОТНИХ КЕПСТРАЛЬНИХ КОЕФІЦІЄНТІВ

Кузнєцов Є.В.

Київський національний університет імені Тараса Шевченка, студент

I. Вступ

Незважаючи на бурхливий розвиток обчислювальної техніки в останні роки, задача розпізнавання мови досі не може вважатися повністю розв'язаною. Більш того, її актуальність з плином часу лише збільшується. З зростанням складності приладів зростає і складність керування ними. У зв'язку з цим виникає потреба у використанні простіших методів керування. Одним із таких методів є голосове керування.

Для розпізнавання людської мови зазвичай використовують:

- динамічне програмування – часові динамічні алгоритми (Dynamic Time Warping);

- методи дискримінантного аналізу, основані на Бассовській дискримінації (Bayesian discrimination);

- приховані Марківські моделі (Hidden Markov Model);
- нейронні мережі (Neural Networks).

У якості ознак для розпізнавання людської мови можна використовувати:

- спектрально-часові ознаки;
- кепстральні ознаки;
- амплітудно-частотні ознаки;
- ознаки нелінійної динаміки.

II. Опис системи

Задачею цієї системи є віднесення звуку на будь-якому малому проміжку сигналу до деякого класу звуків. У якості класів звуків були взяті фонемі російської мови, їх 43.

Система складається з препроцесору та мультикласифікатору. Далі коротко наведено принцип роботи системи:

1. Препроцесор відділяє корисний сигнал від шумів та тиші за амплітудними показниками.
2. Препроцесор розбиває корисний сигнал на ділянки - фрейми (довжини 25 мс), що перекриваються (фактор перекриття – 0.5).
3. Обчислюються мел-частотні кепстральні коефіцієнти[2] для кожного фрейму (13 шт).
4. На вхід мультикласифікатору надходять мел-частотні кепстральні коефіцієнти деякого фрейму, а той, в свою чергу, надає для цього фрейму ймовірності відповідності різним класам.
5. Крок 4 повторюється для кожного фрейму.

Мультикласифікатором є система бінарних класифікаторів, що використовують Error Correcting Output Codes (ECOC) [4]. Саме такий підхід дозволяє використовуючи бінарні класифікатори отримати розподіл ймовірностей належності до класів. Для побудови ECOC-матриці був використаний Randomized Hill Climbing Algorithm [1] з кількістю класів – 43 та кількістю бінарних класифікаторів – 473. У якості бінарних класифікаторів використовувались перцептрони [3] з одним прихованим шаром з кількістю нейронів ($2 \cdot \text{кількість ознак для розпізнавання} = 26$).

III. Навчання

Навчання мультикласифікатору проводилося на фонетично розміченій програмою Festival базі звукових файлів, записаних за участю російськомовного диктору. Кожен звуковий файл має такі параметри: частота – 16 kHz, біт-рейт – 16 kbps. Після обробки препроцесором цих файлів була отримана вибірка розміром 389238 екземплярів. Кожен екземпляр містив 13 мел-частотних кепстральних коефіцієнтів та номер класу якому належить. Для навчання випадковим чином було обрано 19000 екземплярів. Це, в свою чергу, дозволило отримати збалансовану навчальну вибірку.

IV. Результати

Точність системи на тренувальній вибірці досягла 89,37%, в той час як на тестовій – лише 83,21%. Враховуючи те, що існують дуже схожі звуки (наприклад, «д» та «т» схожі), що часто плутаються навіть людиною, отримана точність може вважатися високою. Тестування показало, що переважна більшість помилок – звуки, розпізнані як схожі на них.

Висновки

Була побудована система розпізнавання звуків людської мови на основі мел-частотних кепстральних коефіцієнтів та нейронних мереж. Фактично, ця система є повноцінною акустичною моделлю і може разом з мовною моделлю бути використаною для побудови повноцінної системи розпізнавання людської мови.

Враховуючи використання ECOC, дана система є надзвичайно перспективною. У подальшому планується створення на основі цієї системи повноцінної системи розпізнавання людської мови для української та російської мов.

Список використаних джерел

1. Thomas G. Dietterich and Ghulum Bakiri. Solving multiclass learning problems via error-correcting output codes. Journal of Artificial Intelligence Research, 2:263–286, January 1995.
2. Cheong Soo Yee and Abdul Manan Ahmad, Malay Language Text Independent Speaker Verification using NN-MLP classifier with MFCC, 2008 International Conference on Electronic Design.
3. Simon S. Haykin. Neural Networks: A Comprehensive Foundation, 1999.
4. Robert E. Schapire. Using output codes to boost multiclass learning problems. Machine Learning: Proceedings of the Fourteenth International Conference, 1997.